REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
29-Jun-2011	Phase 1 Final Report	Jan - Jun 2011

4. TITLE AND SUBTITLE

WIPCast:
Probabilistic Forecasting for Aviation Decision Aid Applications

5a. CONTRACT NUMBER
W911QX-11-C-0005

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Hyam Singer, Impact Computing Corporation

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Impact Computing Corporation, 606 Burnt Mills Avenue, Silver Spring, MD 20901
University of Washington, Department of Atmospheric Sciences, ATG Building, Seattle, WA 98195-1640

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

United States Army Research Laboratory
RDRL-CIE-M (Richard Szymber)
WSMR, NM 88002-5513

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
ARL-CR-990

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A fundamental shift in weather forecasting – to an approach that embraces probabilistic forecasting and forecast uncertainty – is crucial to advancing the science and utility of weather prediction. Toward that end, Impact Computing has teamed with the University of Washington's Department of Atmospheric Sciences to develop Weather Impact Probability Forecast (WIPCast) – a solution that builds upon the seminal work of Dr. Tony Eckel in the domain of probabilistic forecasting and its practical application.
Our Phase I effort has laid the foundation for developing a decision support tool that substantially automates the risk analysis process via informed risk analysis whereby a user evaluates projected weather-related risk relative to their risk tolerance. The Phase I prototype in particular has yielded promising results, demonstrating our ability to effectively:

- Use high resolution, mesoscale ensemble model outputs to produce forecast probabilities of meteorological variables critical to army aviation
- Use calculated forecast probabilities to determine anticipated weather impacts on flight routes, in support of operational risk assessment and management
- Provide an intuitive display of overall adverse weather impact probabilities along a flight path

The importance of this capability to U.S. military operations cannot be overstated. In the commercial marketplace as well, a wide array of weather-sensitive businesses are critically dependent upon their ability to anticipate and manage weather-related risk.

15. SUBJECT TERMS

probabilistic forecasts, ensemble forecasts, weather, decision support, risk analysis, uncertainty analysis

16. SECURITY CLASSIFICATION OF: UNCLASSIFIED			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Hyam Singer
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	45	**19b. TELEPHONE NUMBER** *(include area code)* 301-593-2350 x701

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

ARMY RESEARCH LABORATORY

WIPCast

Probabilistic Forecasting for Aviation Decision Aid Applications

SBIR Topic A10-065 Phase I Final Report

ARL-CR-990 **June 2011**

<u>Prepared by:</u>

Impact Computing Corporation
606 Burnt Mills Avenue, Silver Spring, MD 20901

University of Washington
Department of Atmospheric Sciences

Under Contract: W911QX-11-C-0005

United States Army Research Laboratory

RDRL-CIE-M (Richard Szymber)

WSMR, NM 88002-5513

Approved for public release; distribution is unlimited.

Document Revision 1.0
29-Jun-2011

Table of Contents

Table of Contents
(continued)

1 EXECUTIVE SUMMARY

A fundamental shift in weather forecasting – from an approach that is predominantly deterministic to one that embraces communication of forecast uncertainty (i.e., probability-based) – is crucial to advancing the science and application of weather prediction (NRC 2006). Probabilistic forecasts provide a critical added dimension of information that quantifies risk (i.e., the probability of undesirable negative outcomes) and enables optimal decision making by users. While there are many approaches to generating probabilistic forecasts (e.g., based on statistical analysis of past forecast errors), the most promising technique today is the combination of ensemble forecasting with statistical post-processing. Ensemble forecasts dynamically simulate the flow-dependent forecast uncertainty by varying initial conditions and model processes across multiple runs of a numerical weather prediction model.

Probability forecasts with high reliability and resolution can be used effectively in decision making processes to manage weather-related risk (Eckel et al. 2008). The importance of this capability to the success of U.S. military operations cannot be overstated. In the commercial marketplace as well, a wide array of weather-sensitive businesses are critically dependent upon their ability to effectively anticipate and manage weather-related risk.

Accordingly, the value of probabilistic forecasting ultimately lies in its ability to help users – whether they are military planners, commodity traders, or families planning an outing – manage weather-related risk. By quantifying risk, probabilistic forecasting enables optimization of actions via informed risk analysis, in which a user evaluates the projected risks relative to their risk tolerance (i.e., their ability to absorb the impact of the undesirable outcome). To the extent that the user's risk tolerance and the potential outcomes can be characterized and quantified, these factors can be combined with probabilistic forecasts to yield a decision support tool that substantially automates the risk analysis process.

The goal of the Phase I R&D effort described herein has been to lay a firm foundation for the development of such a tool. Toward that end, Impact Computing has teamed with the University of Washington's Department of Atmospheric Sciences to develop an innovative ***Weather Impact Probability Forecast* (WIPCast)** solution that implements and builds upon the seminal work of Dr. Tony Eckel in the domain of probabilistic forecasting and its practical application.[1]

The Phase I effort has yielded promising results that validate the soundness and utility of our approach. In particular, our working Phase I proof-of-concept prototype has demonstrated our ability to effectively:

- Use high resolution, mesoscale ensemble model outputs to produce forecast probabilities of meteorological variables critical to aviation along a particular flight path[2]
- Use calculated forecast probabilities to determine anticipated weather impacts on manned and unmanned aircraft flight routes (for a grid point, flight leg, and flight path) in support of operational risk assessment and management
- Provide an intuitive display of overall adverse weather impact calculations and probabilities at a grid point, along a flight leg, and along an entire flight path

This final report is divided into the following sections:

1. **The WIPCast Solution** (Section 2). An updated technical description of our WIPCast methodology and solution, based on the results of our Phase I effort.
2. **Phase I Review** (Section 3). Details of our Phase I activities, results and conclusions.
3. **Phase I Option** (Section 4). Our plan for the Phase I Option period.
4. **Phase II Plan** (Section 5). Our plan for a successful follow-on Phase II initiative.

[1] Dr. Eckel has been a close and active advisor to our team throughout the Phase I effort.
[2] The Phase I prototype has specifically focused on icing and turbulence.

2 THE WIPCAST SOLUTION

The technical description provided herein describes our *WIPCast* solution as it has evolved over the course of our Phase I effort.

Our innovative approach to the calculation of *Weather Impact Probability (WIP)* is at the core of our *WIPCast* solution. *WIPCast* integrates the rich informational content of probabilistic forecasts with mission impact analysis so as to effectively inform risk analysis (as a function of risk tolerance) to drive an effective and intuitive decision support tool for the end user. In particular, to the extent that the user's risk tolerance can be characterized and quantified, it can be combined with the risk outputs from probabilistic forecasts to yield a decision support tool that substantially automates the risk analysis process.

2.1 KEY INNOVATIONS AND FEATURES

Key innovations and features of our *WIPCast* solution include:

- Calculation of **Weather Impact Probability (*WIP*)** to convey the total risk of serious degradation to the complete mission from weather (see Section 2.3)
- Development and application of **Mission Impact Functions (MIFs)** that describe the uncertainty in the chance of mission failure from actual occurrence of adverse weather (see Section 2.4)
- Objective **calculation of ambiguity** in the ensemble forecast and translation into a *WIP* confidence interval to convey confidence in the decision input (see Section 2.5)
- Statistical amalgamation of the potential impact from multivariate weather sensitivities, based on a **multivariate space-time probability model** for the weather elements of interest, conditioned on the calibrated ensemble forecasts (see Section 2.6)
- **Intuitive visualization and display** of anticipated mission impacts, provided in the context of a comprehensive weather-based mission planning decision support tool (see Section 3.2.6)
- Flexible **Service Oriented Architectures (SOA)** that maximizes interoperability with existing systems and frameworks (see Section 5.3.11)
- **Data source agnostic approach** that avoids reliance on the peculiarities or idiosyncrasies of any particular mesoscale ensemble data source (see Section 5.3.11)

2.2 LIMITATIONS OF CURRENT STATE-OF-THE-ART

To date, the common method for incorporating ensemble forecast data into risk analysis has been to calculate the *threshold probability*, or chance of occurrence for a specific *event*, which is described by a weather element and a critical operating threshold. This approach assumes that there will be serious degradation to the mission if the weather occurs at or above that threshold, purely as a step function. The Tri-Service Integrated Weather Effects Decision Aid (T-IWEDA) contains an extensive database of such thresholds, associated with specific events. The critical threshold is defined as "the point where the occurrence of a meteorological element causes a significant (moderate or severe) impact on a military operation, system, subsystem, or personnel" (Shirkey and Gouveia, 2002). An example of such a threshold is surface winds ≥ 13 kt in a drop zone, where 13 kt is considered critical for a static line airdrop, since winds that fast or higher can result in serious injury to personnel and thus overall mission failure.

Ensemble data (when calibrated) can be used to accurately compute the *threshold probability* [denoted as Pr(wind speed ≥ 13 kt)]. This can be computed as simply the relative frequency of forecasts that exceed the critical threshold, or by using more complex methods such as the "rank method" described in Eckel and Mass 2005.

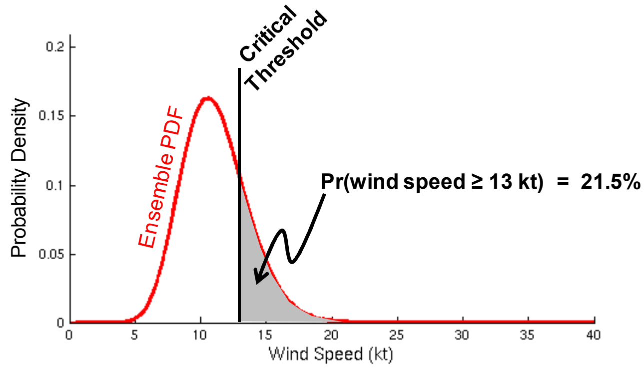


Figure 1. Depiction of the threshold probability, which is the weather risk to a static line airdrop mission for an example ensemble forecast PDF.

Conceptually, all such methods estimate the area under a continuous PDF[3] (fit to the discrete ensemble forecast data) to the right of the critical threshold (Figure 1.). Probability of exceeding the critical threshold of 13 kt is found by integrating (i.e., summing up) the area under the ensemble PDF to the right of the threshold. This type of calculation must be approximated from the discrete ensemble data since the continuous PDF is normally not known. The threshold probability can then be compared to the user's risk tolerance to arrive at an appropriate decision input.

While the approach of comparing threshold probability to risk tolerance can provide valuable input to the decision process, it has the following significant limitations:

- **Limitation #1: Use of fixed, deterministic thresholds**. The operating thresholds in T-IWEDA are designed to be applied to the type of deterministic (i.e., single valued) forecasts that have traditionally been supplied to DoD operators. The threshold is compared to the deterministic forecast and a stoplight chart decision input is presented to the operator. When the forecast exceeds the critical threshold, the decision input is red, meaning "either a total or severe degradation, or the operational limits or safety criteria have been exceeded" (Shirkey and Gouveia, 2002). Since the forecast carries a well understood average amount of uncertainty (i.e., the long-term average error), the T-IWEDA thresholds were likely defined to roughly account for that uncertainty. Continuing with the above example, the 13 kt threshold was chosen (hypothetically) with the knowledge that the forecast has an average error of perhaps 4-5 kt. The true threshold of concern where injuries really begin to mount may actually be more like 17 kt, so in order to hedge on the side of caution given the likely magnitude of forecast error, the critical threshold was set at 13 kt. While that supposition could be argued against, it is clear that fixed, deterministic thresholds have a questionable meaning in the context of ensemble forecasting where, instead of the average uncertainty, the dynamic (or case-by-case) forecast uncertainty is now known and conveyed with a forecast PDF (Figure 1.).

  A revised methodology is required to make best use of ensemble data in calculating the risk by introducing the concept of the impact itself being uncertain even when the weather conditions are known precisely.

  The WIPCast Solution. *WIPCast addresses the problematic application of T-IWEDA thresholds with ensemble data by introducing the concept of a Mission Impact Function (MIF), which describes the impact from weather conditions in probabilistic terms as opposed to fixed, deterministic thresholds. See Section 2.4.*

- **Limitation #2: Use of marginal operating thresholds**. T-IWEDA defines marginal operating thresholds to calculate marginal risk (i.e., yellow/amber stoplight) as well as negligible risk (i.e., green

3 PDF (Probability Density Function) is a function that describes the relative likelihood of a particular continuous random variable occurring at a given point.

stoplight). In our example, 9 kt is the marginal threshold so 9 kt ≤ wind < 13 kt is the marginal operating range and wind < 9 kt corresponds to "no operational restrictions" (Shirkey and Gouveia, 2002). This threshold is inappropriate to apply to ensemble data for the same reasons as described above, but with an additional complication. A yellow decision input *should* be interpreted as unclear risk, whereas green and red mean a clearly acceptable or unacceptable risk, respectively.

In applying ensemble data to calculation of risk, segregation of clear vs. unclear decision inputs requires introducing another new concept called ambiguity. Ambiguity, which refers to the uncertainty in the probability forecast, provides a much more meaningful mechanism for determining the existence of marginal risk.

The WIPCast Solution. *To account for ambiguity, WIPCast calculates the ambiguity in the acquired ensemble forecasts following the methods described by Eckel et al. (2011) and applies that information to calculate confidence intervals for WIP. See Section 2.5.*

- **<u>Limitation #3: Oversimplified handling of multivariate events</u>**. A mission may be impacted by multiple events involving multiple weather elements with various operational sensitivities and different relative importance. Simply applying risk analysis on each event separately can lead to suboptimal, or potentially even erroneous, decisions.

  A novel approach is required that effectively accounts for the correlation between disparate weather elements so that their associated risks can be intelligently analyzed simultaneously.

  The WIPCast Solution. *WIPCast employs an innovative statistical approach to addressing the complex multivariate challenge as described in Section 2.6.*

- **<u>Limitation #4: Graphical representation of risk as a step function</u>**. Risk is not a step function. Solely using thresholds to make a binary (or more precisely trinary) determination – i.e., as to when risk goes from low to moderate to high -- is often inadequate, or potentially even misleading, in the context of an operational decision support tool. The user needs to be able to understand "how red is red" in order to make a fully informed decision as to whether or not to take an educated risk.

  It is crucial for decision support tools used for risk analysis to depict the severity of risk, not solely as a step function, but along more of a continuum. Users need to be able to understand how far beyond a moderate or high risk threshold their anticipated risk lies.

  The WIPCast Solution. *WIPCast provides multiple integrated risk displays that jointly provide the user with deeper insight into their risk, helping them to more precisely quantify the extent of the anticipated risk. See Section 3.2.6.*

2.3 WEATHER IMPACT PROBABILITY (WIP)

The basic concept of a probabilistic impact, and how to apply it in calculating *WIP*, was adapted from a visionary Air Force publication, AWS/TR--91/001, that was unfortunately never adopted into DoD weather support operations (AWS 1991). The approach described therein was to calculate a "mission success indicator" by integrating the product of the forecast PDF and a function describing the chance of no impact. In our case, rather than focusing on the chance of mission success, we focus on risk (or chance of failure), using what we refer to as the Mission Impact Function (MIF) which is simply the inverse of their no-impact function. MIF describes the impact from weather conditions in probabilistic terms as opposed to fixed, deterministic thresholds. (Our approach to calculating MIF is further described in Section 2.4.)

WIP may thus be calculated as:

$$WIP = \int_x \mathrm{F}(x)\ \mathrm{MIF}(x)\ dx$$

where x is the continuous variable (e.g., wind speed) and F is the forecast PDF.

Three sample forecasts and *WIP* results are shown in Figure 2. . The trivial cases are when the range of possible wind speed (described by the ensemble forecast PDF) is so low that no impact is possible and *WIP* = 0% (Figure 2. a), or so high that impact is certain and *WIP* = 100% (Figure 2. c). A threshold probability gives the same risk estimate for these trivial cases. **The interesting case that shows the value of *WIP* is when there is overlap between the two distributions and *WIP* is able to properly convey the overall uncertainty (Figure 2. b).** Comparing Figure 2. b to Figure 1. reveals the limitations of using threshold probability to describe risk. Essentially, risk in Figure 1. is calculated the same way as for *WIP*, but in the place of MIF is a step function (0 for $x < 13$, and 1 for $x \geq 13$) to represent the impact. The resulting threshold probability in this example presents a risk = 21.5% vs. the more robust risk from *WIP* of 30.2%. Threshold probability tends to bias risk toward the extremes (0% and 100%) by underestimating low risk values, when the critical threshold falls on the right side of the forecast PDF, and overestimating high risk values, when the threshold falls on the left side. This bias can notably degrade the risk analysis and the user's performance.

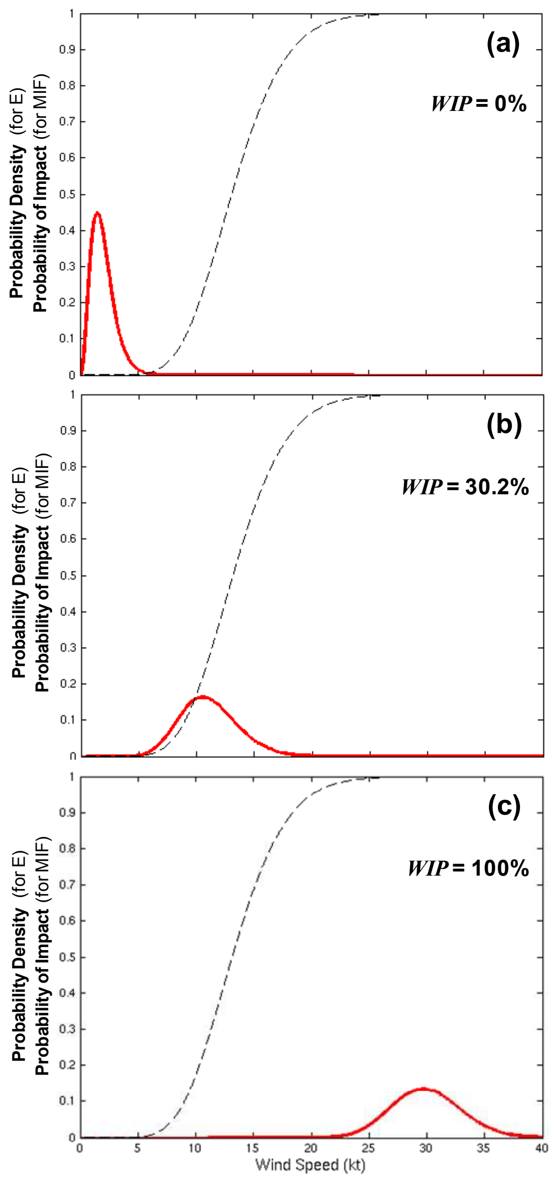


Figure 2. Sample calculations of WIP with the static airdrop MIF (dashed) for three different ensemble forecast PDFs (solid red) in (a) very light wind, (b) a moderately windy, and (c) high wind situations.

A further complicating factor, effectively addressed in our approach to calculating WIP, is that of *ambiguity*. In particular, while the MIF calculation described in Section 2.4 properly represents the risk of mission impact, it is actually an expected value (or best-estimate) of the risk since there is also uncertainty about the risk itself. That uncertainty comes from the fact that a well-calibrated ensemble forecast PDF still contains random error (Eckel et al. 2010) which carries over into *WIP*. This type of uncertainty is termed *ambiguity* since it is 2nd order uncertainty as opposed to an ensemble's 1st order estimate in the forecast uncertainty (NRC 2006).

To account for ambiguity, *WIPCast* calculates the ambiguity in the acquired ensemble forecasts following the methods described by Eckel et al. (2011) and applies that information to calculate confidence intervals for *WIP*. (The details of our approach to calculating ambiguity are further discussed in Section 2.5.) Comparing the *WIP* confidence interval to the user's risk tolerance helps drive highly effective and innovative decision support mechanisms for visually quantifying risk for the user.

2.4 MISSION IMPACT FUNCTION (MIF)

WIPCast addresses the problematic application of T-IWEDA thresholds with ensemble data by introducing the concept of a *Mission Impact Function* (MIF), which describes the impact from weather conditions in probabilistic terms as opposed to fixed, deterministic thresholds.

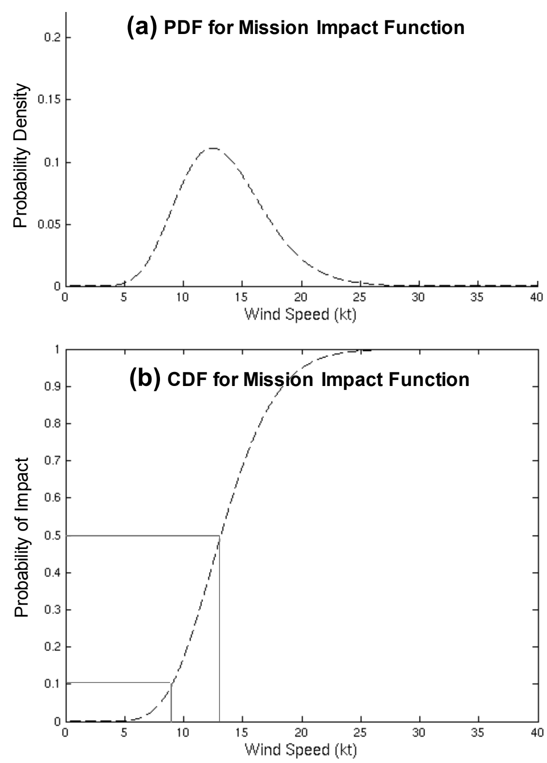


Figure 3. Example MIF, representing the chance of impact to static line airdrop mission for a given observed wind speed as a (a) PDF or (b) CDF. Solid lines in (b) indicate marginal (9 kt) and critical (13 kt) thresholds, tied to the 10th and 50th percentile values respectively, help define MIF.

Returning to the static line airdrop mission example discussed in Section 2.2, there is a chance of impact[4] when the observed wind is 13 kt, but a greater chance of impact at 15 kt and lower (but non-negligible) chance at 10 kt. This continuum is described by a MIF, which can be viewed as a probability distribution function (PDF) (Figure 3. a) but is actually applied in the form of a cumulative distribution function (CDF) (Figure 3. b). In this example, the probability of impact increases rapidly for an observed wind from 5 kt to

[4] "Impact" as used herein refers to a degradation of the mission serious enough to cause mission failure.

13 kt and continues to rise at a slowing rate up to 25 kt. For wind speeds above 25 kt, an impact is guaranteed, and for wind < 5 kt there is no chance of impact. (Note: This sample MIF is only an initial attempt with an unknown degree of fidelity, formulated as described below.)

In the context of military systems, the uncertainty described by a MIF comes from various random influences on weapon system performance. For our example, the uncertainty may come from factors such as the variability of the terrain and its effects on landing, airborne personnel readiness (jumpers' skill, experience, amount of fatigue, etc.), and accuracy of the insertion. The way to interpret a MIF of 75%, for example, is that in 3 out of 4 drops with the same conditions (16 kt winds, same landing zone, etc.), those random factors cause a sufficient number of injuries for the mission to fail. However, in 1 out of those 4 drops the mission will "get lucky" and suffer few enough injuries to proceed. Note that MIF should <u>not</u> be interpreted as a percent, or fractional degradation to the mission.

2.5 CALCULATING AMBIGUITY

The primary causes of ambiguity are (a) limited sampling by an ensemble, due to the discrete number of members, and (b) deficiencies in the ensemble's ability to simulate all sources of forecast uncertainty (Eckel et al. 2011). The effect of limited sampling alone can be very significant for a small ensemble, as displayed in Figure 4. (Allen 2009).

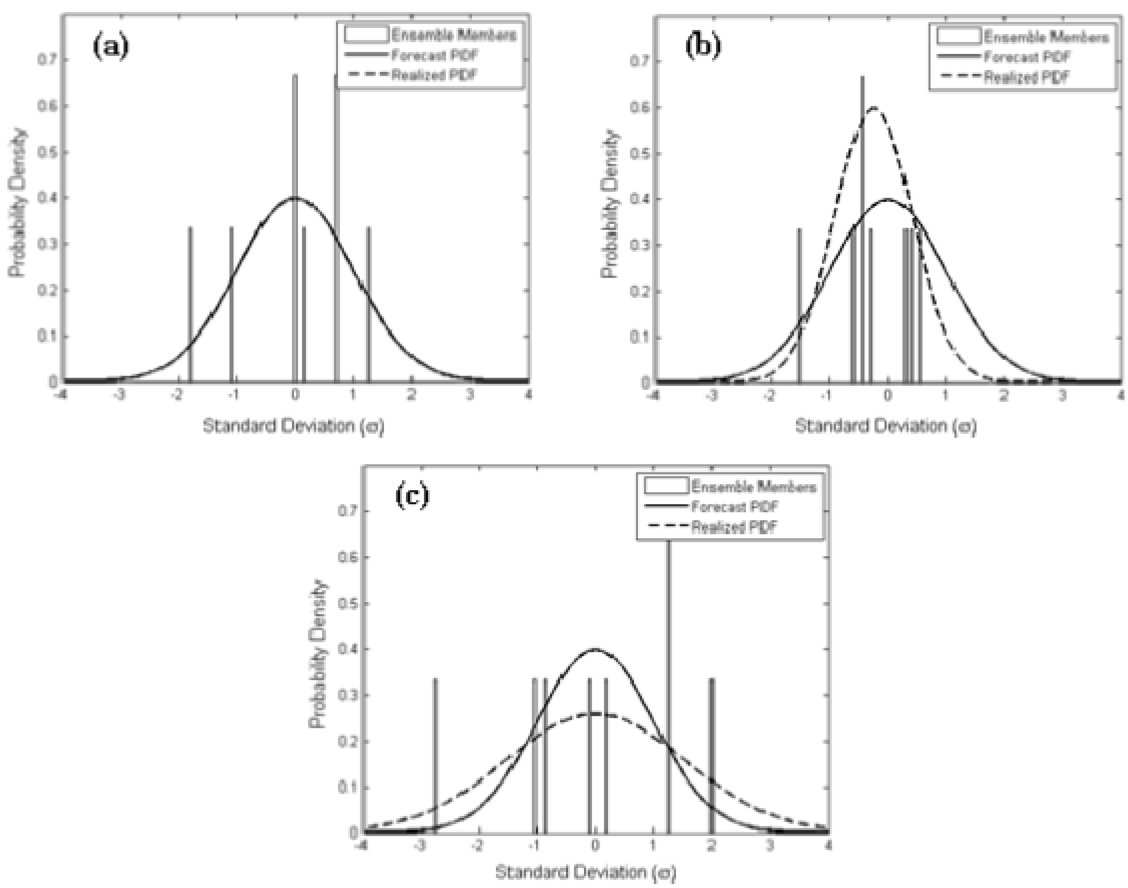


Figure 4. (Taken from figure 1, Allen 2009) Demonstration of ambiguity due to limited ensemble sampling. An 8-member perfect ensemble was simulated by sampling from a hypothetical true forecast PDF (solid line), defined as N(0,1). Ensemble members are represented with a histogram; realized, ensemble PDF is the dashed line. (a) Ensemble correctly represents the true PDF. (b) Ensemble PDF has a random error centered too low with too small a spread. (c) Ensemble PDF has a random error with too high a spread.

A perfect ensemble that samples from the true PDF will be statistically consistent over many cases (i.e., perfectly reliable), but can err notably in representing the true PDF on any one case and thus err in the estimate of forecast probability (or risk) for a specific event. Given an imperfect ensemble, as in today's operational ensemble systems, the problem is amplified since sensitivities to sources of forecast uncertainty poorly simulated in the ensemble can vary case-by-case, thus compounding the random error in the ensemble forecast PDF (Eckel et al. 2011).

Eckel et al. (2011) developed statistical methods to estimate ambiguity and applied them to operational ensemble forecasts. Figure 5. shows that ambiguity can be quite significant even for a large, 51-member ensemble. It also shows that a sharper forecast PDF (i.e., smaller spread) is more vulnerable to ambiguity effects since a small shift of the ensemble's PDF relative to the true PDF can cause a large error in the ensemble's forecast probability estimate (Eckel et al. 2011).

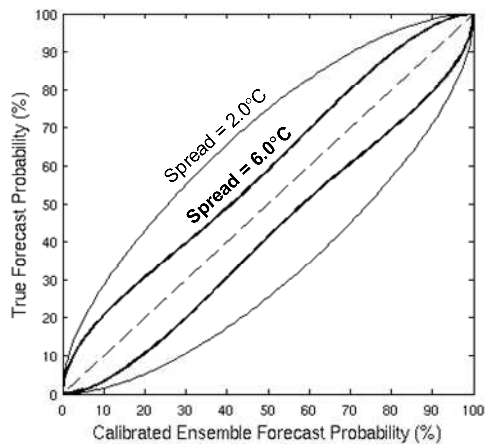


Figure 5. (Taken from figure 13, Eckel et al. 2011) Ambiguity represented as a 90% confidence interval about calibrated, 5-day, 2-m temperature forecasts from the 51-member Japanese Meteorological Agency operational ensemble forecast. The dashed line is the expected value from the calibrated forecasts, which demonstrate perfect reliability. The solid lines show the upper and lower bound of the range of possible true value of probability given an ensemble spread of 2.0°C (thin line) and 6.0°C (thick line).

Szczes (2008) showed how ambiguity information is critical to properly defining when a risk is clearly acceptable (green stoplight) or unacceptable (red stoplight) vs. unclear (yellow stoplight). Basically, when the confidence interval (i.e., range of possible values) about the expected value of risk overlaps the user's risk tolerance, the decision to take action is unclear since the true risk could go either way. Allen and Eckel (2011) further demonstrated that ambiguity information can be applied in the decision process to benefit the user. This involves consideration of secondary factors important to the user that are not easily incorporated into the risk tolerance, such as personnel morale. The idea is that when the decision to act is unclear, the user may base the decision on those secondary factors, thus maintaining the overall optimization of the primary concerns while also optimizing with respect to the secondary ones.

From a decision support standpoint, a key outcome of our Phase I effort is the redefining of marginal risk as impact values within the range of ambiguity in the risk (i.e., the risk "spread"), as shown in Figure 6. .

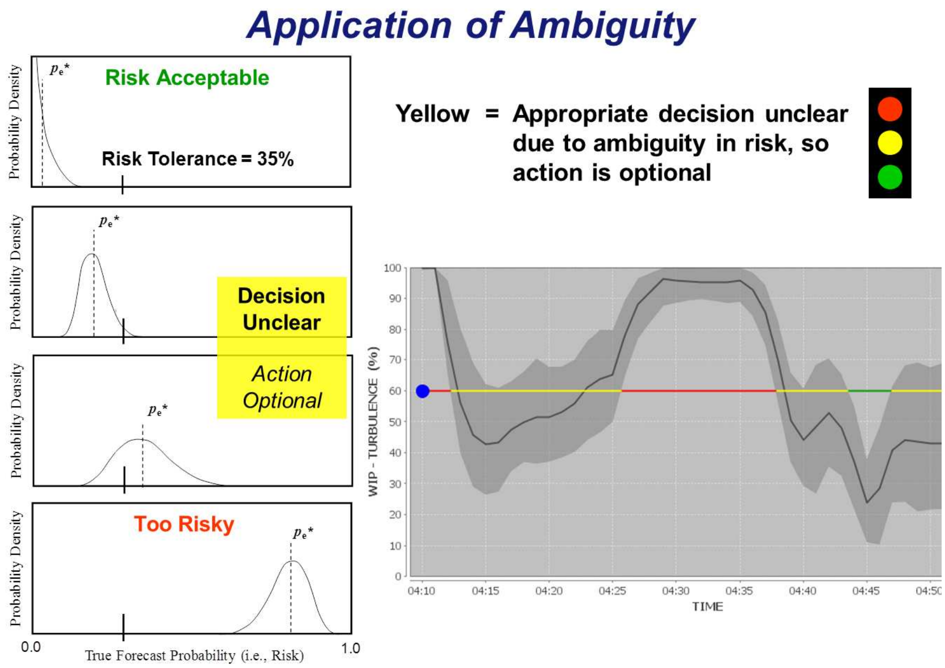


Figure 6. Redefining marginal risk (what does "yellow" mean?)

2.6 MULTIVARIATE RISK CALCULATION (TOTAL WIP)

So far, we have limited our discussion to scenarios where *WIP* depends on a single weather element. In practice, however, the chance of an impact (and mission failure) depends on multiple elements, such as wind speed and direction, visibility, turbulence, and icing. In addition, it depends on these quantities at multiple points in both time and space, for example, over an entire flight path. Thus, in practice, the total *WIP* depends on a large number of inter-related weather phenomena simultaneously.

Although meeting the challenge of accurately calculating a total *WIP* that accounts for concurrent risks from multiple weather sensitivities has been deferred to Phase II, a proof-of-concept simplified approach was employed in Phase I. In the Phase I prototype, WIP was generated separately for two weather sensitivities (turbulence and icing) and a total WIP was produced that accounts for both risks using a simplification. Further detail is provided in Section 5.3.3.

2.7 CALIBRATION OF PROBABILISTIC FORECASTS

Today, the resolution of ensemble systems has increased sufficiently for mesoscale circulations to be predicted probabilistically. However, the output from mesoscale ensemble systems should not be used directly in the production of probabilistic forecasting guidance for aviation or other uses. All forecasting systems possess systematic bias that does not represent forecast uncertainty, much of which can be removed by post-processing (Eckel and Mass, 2005). *Post-processing of ensemble member output can potentially greatly improve the reliability and sharpness of the final probabilistic products derived from ensemble systems.* Post-processing can range from simple bias correction (e.g., Baars and Mass, 2005) to sophisticated Bayesian Model Averaging (BMA) approaches that include bias correction, weighting of ensembles based on past performance, and adjusting the variability of each member to produce optimal probability density functions (Raferty et al., 2005).

WIPCast needs an effective, efficient calibration approach that easily adapts to the wide range of weather variables of importance to aviation. Many of the more sophisticated techniques do not meet these criteria. *WIPCast* Phase I therefore employed the simple *shift-and-stretch* calibration technique described in Eckel and Allen (2011). The 1st moment bias is defined by the mean error (*ME*) in the ensemble mean:

$$ME_{\bar{e}} = \frac{1}{M}\sum_{i=1}^{M}\left(\bar{e}_i - o_i\right)$$

where M is the total number of individual verifications (indexed by i), $\bar{e}_i$ is the ensemble mean and o_i is the observation. Calibration of the 1st moment is thus accomplished by shifting all n members by a *shift factor*:

$$\tilde{e}_i = e_i + shift.factor \qquad \text{for } i = 1 \ldots n$$

$$shift.factor = -ME_{\bar{e}}$$

The 2nd moment bias is defined as a fractional error following the concept of statistical consistency. For a consistent ensemble, the mean square error of the bias-corrected ensemble mean ($MSE_{\bar{e}}$) matches the average ensemble variance ($\overline{\sigma_e^2}$) over a large verification dataset, so that $\overline{\sigma_e^2} / MSE_{\bar{e}} = 1$.

$$MSE_{\bar{e}} = \left(\frac{n}{n+1}\right)\frac{1}{M}\sum_{i=1}^{M}\left(\bar{e}_i - o_i\right)^2$$

$$\overline{\sigma_e^2} = \frac{1}{M}\sum_{i=1}^{M}\left[\frac{1}{n-1}\sum_{j=1}^{n}\left(e_{i,j} - \bar{e}_i\right)^2\right]$$

where n is the number of members and $e_{i,j}$ is the j^{th} ensemble member (Eckel and Mass 2005). The fractional error (FE) in average ensemble spread is given by:

$$FE = \sqrt{\frac{\overline{\sigma_e^2}}{MSE_{\bar{e}}}}$$

Calibration is thus accomplished by stretching apart (or compressing) the 1st moment bias-corrected members about their mean ($\overline{\tilde{e}}$) by a *stretch factor* to correct for spread that is too small (or large):

$$e_i^* = \overline{\tilde{e}} + \left(\tilde{e}_i - \overline{\tilde{e}}\right) * stretch.factor \qquad \text{for } i = 1 \ldots n$$

$$stretch.factor = \frac{1}{FE}$$

The fully adjusted members (e_i^*) are then used to produce calibrated forecast probability following the *uniform ranks* method. Basically, forecast probability is found by (see Hamill and Colucci 1997 for complete explanation):

1. Identifying the rank-ordered position of the event threshold (e.g. wind speed ≥ 13 kt) within the n + 1 possible locations among the n ensemble members.

2. Summing probability from all ranks above the threshold, where each rank contains probability of 1/(n + 1).

3. Adding in a linear fraction of the probability from the rank in which the event threshold occurs.

Besides producing continuous values of probability, this technique also alleviates biasing toward extreme probability that occurs with the relative frequency technique (Eckel, 2003).

2.8 Weather Observation Data ("Ground Truth")

For training the calibration and ambiguity routines, and for verification studies, *WIPCast* needs high quality weather observations of all forecast variables of interest to aviation. This includes not only surface information such as winds, visibility, cumulative precipitation, etc. for airfield and surface operations, but also variables throughout the volume of the atmosphere such as winds, temperature, and relative humidity. Furthermore, the data needs to continuously cover large operations' domains. A gridded analysis of better or comparable resolution to the forecast data is the best solution, but is unfortunately not always available. The methodology employed for the Phase I proof-of-concept is described in Section 3.2.8 and our planned approach for Phase II is discussed in Section 5.3.8.

3 PHASE I REVIEW

As detailed in the subsections that follow, our Phase I effort has proven highly productive, yielding results that will serve as a firm foundation for a follow-on Phase II initiative.

3.1 PHASE I OBJECTIVES AND ACCOMPLISHMENTS

Our Phase I initiative has focused on development of a proof-of-concept prototype, through which all technical goals for Phase I have been successfully achieved and demonstrated, as listed in 0.

<table>
<tr><th colspan="2">PHASE I OBJECTIVES</th><th rowspan="2">PHASE I ACCOMPLISHMENTS</th></tr>
<tr><th>AS STATED IN SBIR SOLICITATION</th><th>AS STATED IN PHASE I PROPOSAL</th></tr>
<tr><td>Develop approaches and methodologies for using mesoscale ensemble modeling to produce forecast probabilities of meteorological variables critical to aviation.</td><td rowspan="5">Use high resolution, mesoscale ensemble model outputs to produce forecast probabilities of meteorological variables critical to aviation (e.g., such as icing, turbulence, thunderstorms, winds, and clouds) along a particular flight path.
Use calculated forecast probabilities to determine anticipated weather impacts on manned and unmanned aircraft flight routes (for a grid point, flight leg, and flight path) in support of operational risk assessment and management.</td><td>✓</td></tr>
<tr><td>Establish capability to calculate forecast probabilities for meteorological parameters using ensemble model forecasts for determination of weather impacts on manned and unmanned aircraft flight routes.</td><td>✓</td></tr>
<tr><td>Determine technical feasibility and develop approaches to producing probabilistic forecasts of specific aviation weather variables such as icing, turbulence, thunderstorms, winds, and clouds from mesoscale ensemble model output for aviation meteorology.</td><td>✓</td></tr>
<tr><td>Establish a method (or methods) of calculating ensemble model forecast probabilities for aviation meteorological parameters and apply these methods to individual parameter probabilistic predictions.</td><td>✓</td></tr>
<tr><td>Develop an initial capability to take one parameter’s adverse weather threshold probability of occurrence and translate that to an adverse weather impact display for a grid point, flight leg, and an entire flight route.</td><td>✓</td></tr>
<tr><td>Develop an initial capability to take one parameter’s adverse weather threshold probability of occurrence and translate that to an adverse weather impact display for a grid point, flight leg, and an entire flight route.</td><td>Provide an intuitive display of overall adverse weather impact calculations and probabilities at a grid point, along a flight leg, and along an entire flight path.</td><td>✓</td></tr>
</table>

Phase I Objectives vs. Accomplishments

3.2 PHASE I PROTOTYPE

Development of our Phase I prototype has been a highly productive and educational exercise, yielding the following key benefits:

1. Served as a focal point for our discussions with the customer and an interactive mechanism for better communicating and demonstrating some of our key portrayal concepts.
2. Focused our efforts on achieving demonstrable, measurable results.

Our Phase 1 prototyping efforts have yielded a preliminary implementation of the *WIPCast* processing pipeline including:

- Calculation of **Weather Impact Probability (*WIP*)**
- Development and application of a **Mission Impact Functions (MIF)**
- Objective **calculation of ambiguity** in the ensemble forecast and translation into a *WIP* confidence interval to convey confidence in the decision input
- Statistical amalgamation of the potential impact from multivariate weather sensitivities by extending *WIP* calculation to use of a **Joint Predictive PDF** decision input
- ***WIPCast* displays** that intuitively and clearly present anticipated adverse weather impacts and risks at a grid point, along a flight leg, and along an entire flight path

The proof-of-concept implementation focused on icing and turbulence as examples of specific aviation weather variables of interest.

3.2.1 DERIVED WEATHER PARAMETERS

Forecasts for two weather phenomena were considered in Phase I (i.e., en route turbulence and icing) and generated from the raw WRF model output in post-processing algorithms. A *turbulence index* and an *icing index*, derived separately on each ensemble forecast member, were used to represent the intensity of each phenomenon on a continuous scale. Fixed threshold values of the indices were used to demark the standard light (LGT), moderate (MDT), and severe (SVR) intensity levels as applicable to aviation.

The turbulence index (*TI*) algorithm followed the method presented by Ellrod and Knapp (1992), which considers the wind shear, deformation and convergence in the forecasted, three-dimensional model wind field. Given the UWME model resolution of 12km, we followed the Air Force Weather Agency's (AFWA's) empirically based thresholds for mesoscale models of LGT ($3 < TI \leq 9$), MDT ($9 < TI \leq 14$), and SVR ($14 < TI$). This produced good results at upper levels (>10,000ft) but often considerably over-forecasted turbulence in the low levels. This issue can be alleviated by implementing the Panofsky Index (PI) to represent turbulence in the lower levels (Boyle 1990, Passner 2000, Brooks and Oder 2004). There was insufficient time in *WIPCast* Phase I for this implementation, so the Phase I turbulence forecasts have a recognized high bias in the lower levels. Implementing the PI is a task for *WIPCast* Phase II.

The icing index (*IcgI*) algorithm followed the method developed by AFWA as part of the Joint Ensemble Forecast System (JEFS) experiment and demonstration. The *II* value (Figure 7.) is based on an empirical formula that uses temperature (*T*, in Kelvin) and relative humidity (*RH*, in %) at a specific model grid point. The standard thresholds are LGT ($58 < IcgI \leq 83$), MDT ($83 < IcgI \leq 90$), and SVR ($90 < IcgI$).

$$IcgI = \mathrm{TANH}(RH*6 - 4) * \mathrm{TANH}((T - 247)/10)*100$$

	-23	-22	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	Temperature (C)
RH(%)	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	Temperature (K)
0.70	5.7	7.5	9.1	10.6	11.9	13.1	14.1	15.0	15.8	16.5	17.0	17.5	17.9	18.2	18.5	18.7	18.9	19.0	19.2	19.3	19.3	19.4	19.5	19.5	0.0	0.0	
0.71	7.4	9.7	11.8	13.7	15.4	16.9	18.2	19.4	20.4	21.2	21.9	22.5	23.0	23.4	23.8	24.1	24.3	24.5	24.7	24.8	24.9	25.0	25.1	25.2	0.0	0.0	
0.72	9.0	11.8	14.3	16.6	18.7	20.6	22.2	23.6	24.8	25.8	26.7	27.4	28.0	28.5	29.0	29.3	29.6	29.8	30.0	30.2	30.3	30.4	30.5	30.6	0.0	0.0	
0.73	10.6	13.8	16.8	19.5	21.9	24.1	26.0	27.6	29.0	30.2	31.3	32.1	32.8	33.4	33.9	34.3	34.7	35.0	35.2	35.4	35.5	35.7	35.8	35.9	0.0	0.0	
0.74	12.0	15.7	19.1	22.2	25.0	27.5	29.6	31.5	33.1	34.5	35.6	36.6	37.4	38.1	38.7	39.2	39.6	39.9	40.1	40.4	40.5	40.7	40.8	40.9	0.0	0.0	
0.75	13.5	17.6	21.4	24.8	27.9	30.7	33.1	35.2	37.0	38.5	39.8	40.9	41.8	42.6	43.2	43.8	44.2	44.5	44.8	45.1	45.3	45.5	45.6	45.7	0.0	0.0	
0.76	14.8	19.3	23.5	27.3	30.7	33.7	36.4	38.7	40.7	42.3	43.8	45.0	46.0	46.8	47.5	48.1	48.6	49.0	49.3	49.6	49.8	50.0	50.1	50.2	0.0	0.0	
0.77	16.1	20.9	25.5	29.6	33.3	36.6	39.5	42.0	44.1	45.9	47.5	48.8	49.9	50.8	51.6	52.2	52.7	53.1	53.5	53.8	54.0	54.2	54.4	54.5	0.0	0.0	
0.78	17.2	22.5	27.3	31.8	35.7	39.3	42.4	45.0	47.4	49.3	51.0	52.4	53.5	54.5	55.3	56.0	56.6	57.0	57.4	57.7	58.0	58.2	58.4	58.5	0.0	0.0	
0.79	18.3	23.9	29.1	33.8	38.0	41.8	45.1	47.9	50.4	52.4	54.2	55.7	56.9	58.0	58.9	59.6	60.2	60.7	61.1	61.4	61.7	61.9	62.1	62.2	0.0	0.0	
0.80	19.3	25.2	30.7	35.7	40.1	44.1	47.6	50.6	53.2	55.4	57.2	58.8	60.1	61.2	62.1	62.9	63.5	64.0	64.4	64.8	65.1	65.3	65.5	65.7	0.0	0.0	
0.81	20.3	26.5	32.2	37.4	42.1	46.2	49.9	53.0	55.7	58.0	60.0	61.6	63.0	64.2	65.1	65.9	66.6	67.1	67.6	67.9	68.2	68.5	68.7	68.9	0.0	0.0	
0.82	21.1	27.6	33.5	39.0	43.9	48.2	52.0	55.3	58.1	60.5	62.6	64.3	65.7	66.9	67.9	68.7	69.4	70.0	70.4	70.8	71.1	71.4	71.6	71.8	0.0	0.0	
0.83	21.9	28.6	34.8	40.4	45.5	50.0	53.9	57.4	60.3	62.8	64.9	66.7	68.2	69.4	70.4	71.3	72.0	72.6	73.1	73.5	73.8	74.1	74.3	74.5	0.0	0.0	
0.84	22.7	29.6	35.9	41.8	47.0	51.7	55.7	59.2	62.3	64.8	67.0	68.9	70.4	71.7	72.8	73.7	74.4	75.0	75.5	75.9	76.2	76.5	76.7	76.9	0.0	0.0	
0.85	23.3	30.4	37.0	43.0	48.4	53.2	57.3	61.0	64.1	66.7	69.0	70.9	72.5	73.8	74.9	75.8	76.5	77.2	77.7	78.1	78.5	78.7	79.0	79.2	0.0	0.0	
0.86	23.9	31.2	37.9	44.1	49.6	54.5	58.8	62.5	65.7	68.4	70.8	72.7	74.3	75.7	76.8	77.7	78.5	79.2	79.7	80.1	80.5	80.8	81.0	81.2	0.0	0.0	
0.87	24.5	31.9	38.8	45.1	50.7	55.8	60.1	63.9	67.2	70.0	72.4	74.3	76.0	77.4	78.5	79.5	80.3	80.9	81.5	81.9	82.3	82.6	82.8	83.0	0.0	0.0	
0.88	25.0	32.5	39.6	46.0	51.8	56.9	61.3	65.2	68.6	71.4	73.8	75.8	77.5	78.9	80.1	81.1	81.9	82.6	83.1	83.6	83.9	84.3	84.5	84.7	0.0	0.0	
0.89	25.4	33.1	40.3	46.8	52.7	57.9	62.4	66.4	69.8	72.7	75.1	77.2	78.9	80.3	81.5	82.5	83.4	84.0	84.6	85.1	85.4	85.7	86.0	86.2	0.0	0.0	
0.90	25.8	33.6	40.9	47.5	53.5	58.8	63.4	67.4	70.9	73.8	76.3	78.4	80.1	81.6	82.8	83.8	84.7	85.4	85.9	86.4	86.8	87.1	87.4	87.6	0.0	0.0	
0.91	26.1	34.1	41.5	48.2	54.3	59.6	64.3	68.4	71.9	74.8	77.4	79.5	81.3	82.7	84.0	85.0	85.8	86.5	87.1	87.6	88.0	88.3	88.6	88.8	0.0	0.0	
0.92	26.5	34.5	42.0	48.8	54.9	60.3	65.1	69.2	72.7	75.8	78.3	80.5	82.3	83.8	85.0	86.0	86.9	87.6	88.2	88.7	89.1	89.4	89.7	89.9	0.0	0.0	
0.93	26.8	34.9	42.5	49.3	55.5	61.0	65.8	70.0	73.5	76.6	79.2	81.3	83.1	84.7	85.9	87.0	87.8	88.6	89.1	89.6	90.0	90.4	90.6	90.9	0.0	0.0	
0.94	27.0	35.2	42.9	49.8	56.1	61.6	66.4	70.6	74.2	77.3	79.9	82.1	84.0	85.5	86.8	87.8	88.7	89.4	90.0	90.5	90.9	91.2	91.5	91.7	0.0	0.0	
0.95	27.2	35.5	43.2	50.2	56.5	62.1	67.0	71.2	74.9	78.0	80.6	82.8	84.7	86.2	87.5	88.6	89.4	90.2	90.8	91.3	91.7	92.0	92.3	92.5	0.0	0.0	
0.96	27.5	35.8	43.6	50.6	57.0	62.6	67.5	71.8	75.4	78.6	81.2	83.4	85.3	86.9	88.2	89.2	90.1	90.9	91.5	92.0	92.4	92.7	93.0	93.2	0.0	0.0	
0.97	27.6	36.1	43.8	51.0	57.3	63.0	68.0	72.3	76.0	79.1	81.8	84.0	85.9	87.5	88.8	89.8	90.7	91.5	92.1	92.6	93.0	93.3	93.6	93.8	0.0	0.0	
0.98	27.8	36.3	44.1	51.3	57.7	63.4	68.4	72.7	76.4	79.6	82.3	84.5	86.4	88.0	89.3	90.4	91.3	92.0	92.6	93.1	93.5	93.9	94.2	94.4	0.0	0.0	
0.99	28.0	36.5	44.3	51.5	58.0	63.7	68.7	73.1	76.8	80.0	82.7	85.0	86.9	88.4	89.8	90.8	91.8	92.5	93.1	93.6	94.0	94.4	94.7	94.9	0.0	0.0	
1.00	28.1	36.6	44.5	51.8	58.3	64.0	69.1	73.4	77.2	80.4	83.1	85.4	87.3	88.9	90.2	91.3	92.2	92.9	93.6	94.1	94.5	94.8	95.1	95.3	0.0	0.0	

Figure 7. Icing index (IcgI) values with shaded levels of LGT (yellow), MDT (orange), and SVR (red) icing intensity levels.

3.2.2 Ensemble Forecast Calibration

Since *WIPCast* Phase I dealt with derived parameters, there was a choice between calibrating the state variables used in deriving the final parameters (e.g., wind components for *TI*) or directly calibrate the final parameters. We chose the latter to avoid the complications presented by possible correlations in the state variables involved in the *TI* and *IcgI* calculation. Shift and stretch factors for a particular forecast were computed using the previous 30-days of forecasts/observations as the training period in order to identify seasonally dependent systematic errors. The factors were broken out by vertical level (25 pressure levels) since biases can be very different in the boundary layer vs. the free atmosphere. The factors were also broken out by forecast lead time since the ensemble PDF's systematic errors can be highly dependent on the increasing error growth during the forecast cycle.

For *TI*, we ignored the 1st moment correction which was not meaningful to identify in Phase I because of the limitation in *TI* calculation described above. Figure 8. shows the stretch factors for *TI* for two forecast lead times. The results here indicate the ensemble is underspread (stretch factors > 1.0) as is normally the case for an ensemble. Also note the natural finding here that a longer lead time generally has a lower stretch factor as the ensemble PDF gets closer to saturating to climatology.

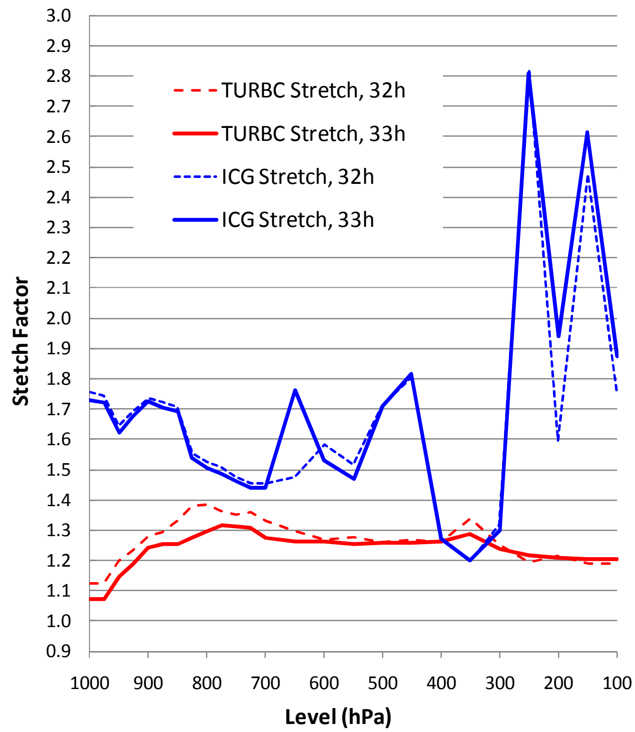


Figure 8. Stretch factors for TURBC and ICG for forecast hours 32 and 33.

For *IcgI*, a significant bias was found which could not be ignored. Furthermore, the bias was found to be heavily dependent on the *IcgI* index itself so we implemented a conditional correction as shown in Figure 9. where the shift factor is a function of *IcgI*. For the ICG (icing) stretch factors, Figure 9. shows a much larger necessary stretching compared to TURBC (turbulence) likely due to the ensemble's difficulty in simulating the range of forecast errors for moisture.

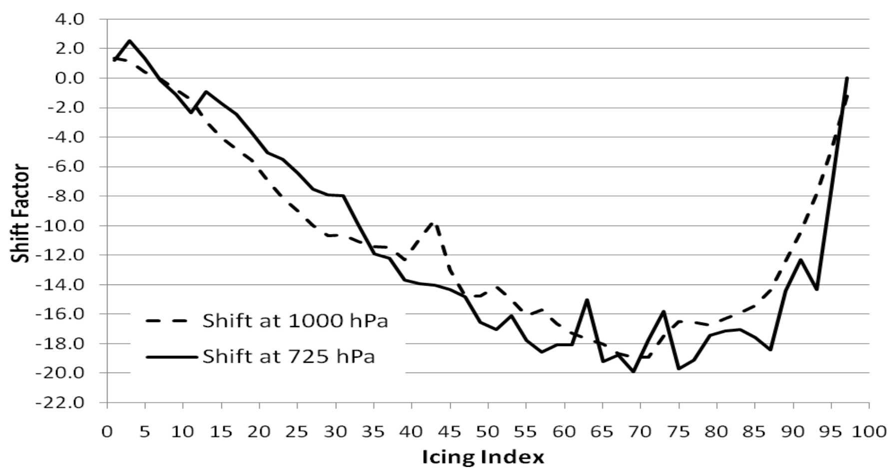


Figure 9. Shift factors for ICG at the 1000 and 725 hPa data levels for forecast hour 32.

There is a notable complication in applying shift-and-stretch to non-Gaussian variables, which we are dealing with in *WIPCast*. *TI* is a positive definite variable (bound by zero on the left) and appears to follow the gamma distribution:

$$f(x;\alpha,\beta) = x^{\alpha-1}\frac{e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \qquad \text{for } x \geq 0 \text{ and } \alpha,\beta > 0$$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

Where *x* is the random variable and α and β are the gamma parameters. *Icgl* values exist between the interval of 0 to 100 so likely follow the beta distribution:

$$f(x;p,q) = \frac{1}{\mathrm{B}(p,q)} x^{p-1}(1-x)^{q-1} \qquad \text{for } 0 \le x \le 1 \text{ and } p,q > 0$$

$$\mathrm{B}(x;p,q) = \int_0^x t^{p-1}(1-t)^{q-1} dt$$

where *p* and *q* are the beta parameters. The problem for shift-and-stretch is that adjusting the two moments of the PDF becomes intertwined since a shift can affect the 2nd moment when accounting for non-physical index values (negative *TI* or 0 > *Icgl* > 100) and likewise a stretch can affect the 1st moment.

The solution is to apply the corrections on $\bar{e}$ and σ_e as if the PDF were Gaussian but find the calibrated member values ($e*$) using the PDF's natural distribution following these steps:

1) Fit the natural PDF to the raw ensemble members using the method of moments
2) Find the original percentile of each ensemble member using the fitted PDF.
3) Fit the natural PDF to the shifted and stretched ensemble mean and standard deviation using the method of moments. (Note: When the ensemble mean falls out-of-bounds, it must be reset to the extreme limit. This does impact the calibration but has no impact on the final forecast probability or *WIP* since it represents a tiny adjustment to an extreme value.)
4) Produce the calibrated members by finding the parameter values on the shifted-and-stretched natural PDF associated with the original percentiles.

We illustrate this process in Figure 10. with the following example ICG forecast at a grid point at the 725 hPa level at the 32h lead time (i.e., these are the nine raw *Icgl* values):

EF = (92.6, 92.9, 93.2, 83.9, 84.8, 90.2, 94.4, 86.1, 93.4)

This EF is fit to a beta PDF by plugging the mean ($\bar{e}$ = 90.2) and spread (σ_e = 4.12) into the equations for the method-of-moments approximation of the beta parameters:

$$p = \frac{(90.2)^2(1-90.2)}{(4.12)^2} - 90.2 = 42.2$$

$$q = \frac{p(1-90.2)}{90.2} = 5.05$$

Next, we find the percentile location of each member on the fitted beta PDF

(0.693, 0.723, 0.752, 0.079, 0.106, 0.453, 0.856, 0.159, 0.770)

Now we find the calibrate beta PDF using the adjusted mean and spread:

shifted mean = 90.2 + (-12.33) = 77.84

shifted spread = 4.12*(1.457) = 6.00

which yields *p* = 36.5, *q* = 10.4 by the same method above. Finally, the values for the calibrated members are found using the original percentiles on the calibrated PDF:

EF calibrated = (81.2, 81.7, 82.2, 69.0, 70.1, 77.5, 84.2, 71.8, 82.5)

These values can then be used to produce calibrated forecast probability. We will also see later that the full calibrated beta PDF will be used in the *WIP* calculation.

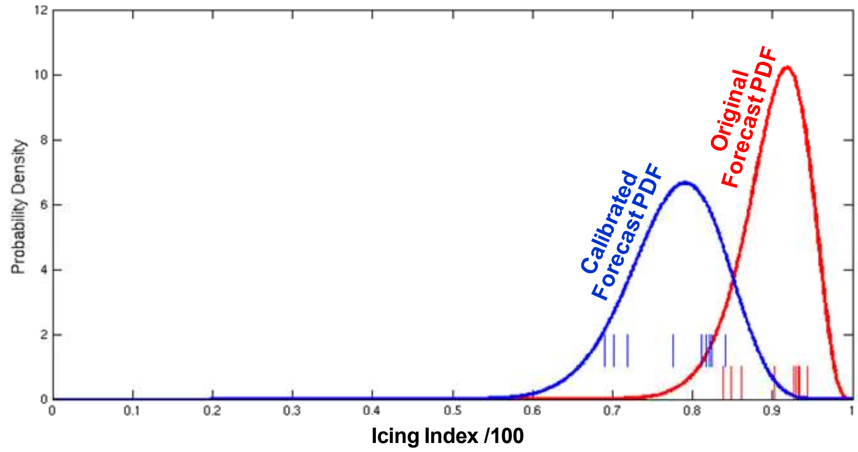


Figure 10. Demonstration of the shift-and-stretch calibration process on a 32-h ICG forecast at 725 hPa. The original nine forecasts of the ensemble data are shown by the lower red lines and the final, calibrated forecasts are the higher, blue lines.

3.2.3 WIP Calculation

WIP was calculated separately for TURBC and for ICG at every model grid point (horizontal and vertical) and lead time using the default MIFs and the calibrated ensemble forecast PDFs. Figure 11. shows a sample result for the risk to the mission from ICG using the example data. Note that the calibrated *WIP* of 94.6% would have been 100% if the uncalibrated forecast PDF (Figure 11.) were applied instead.

For Phase I, we simply assumed that the risk from TURBC is completely independent from the risk from ICG, which provided reasonable results for most cases. The total (or multivariate) *WIP* was therefore found by:

$$WIP_{total} = WIP_{TURBC} + WIP_{ICG} - WIP_{TURBC} * WIP_{ICG}$$

Our planned Phase II approach for more robust calculation of total *WIP* is discussed in Section 5.3.3.

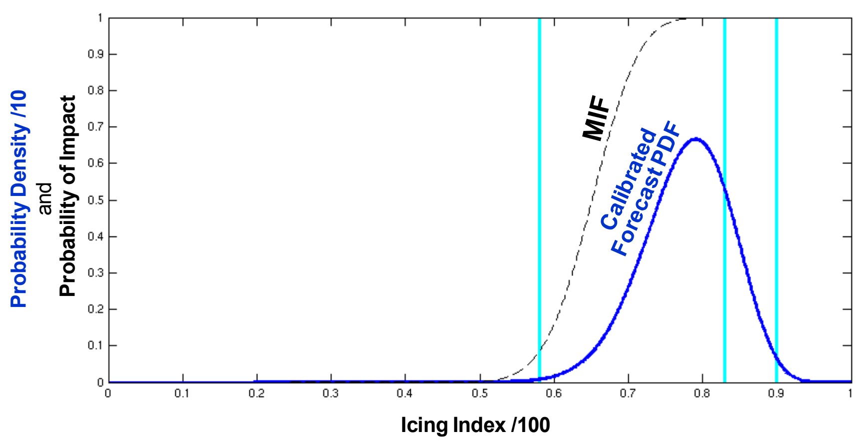


Figure 11. Sample calculation for risk from ICG that yielded a WIP = 94.6%.

3.2.4 Calculating Ambiguity – Randomly Calibrated Resampling

WIPCast output includes 90% confidence intervals about the best-estimate values of forecast probability and WIP based on the ambiguity in the ensemble forecast. For Phase I, we targeted implementation of ambiguity calculation by the randomly calibrated resampling (RCR) method since it is more skillful compared to the calibrated error sampling (CES) method (Eckel and Allen 2011). RCR was successfully coded and run but in an incomplete mode due to the resource constraints of Phase I. Firstly, the random calibration component of RCR was omitted so the ambiguity is based solely on resampling and is thus likely an underestimate. Secondly, performing RCR on the extensive amount of forecast data proved to be extremely costly in computer processing time. To make the *WIPCast* runs manageable, resampling at each grid point for each forecast variable was performed with only 100 samples instead of the desired number of samples on the order of 1,000.

3.2.5 Mission Impact Function (MIF)

Default MIFs for the chance of impact from observed TURBC and ICG intensities were subjectively designed based on likely sensitivities of rotary-wing aircraft (Figure 12.).

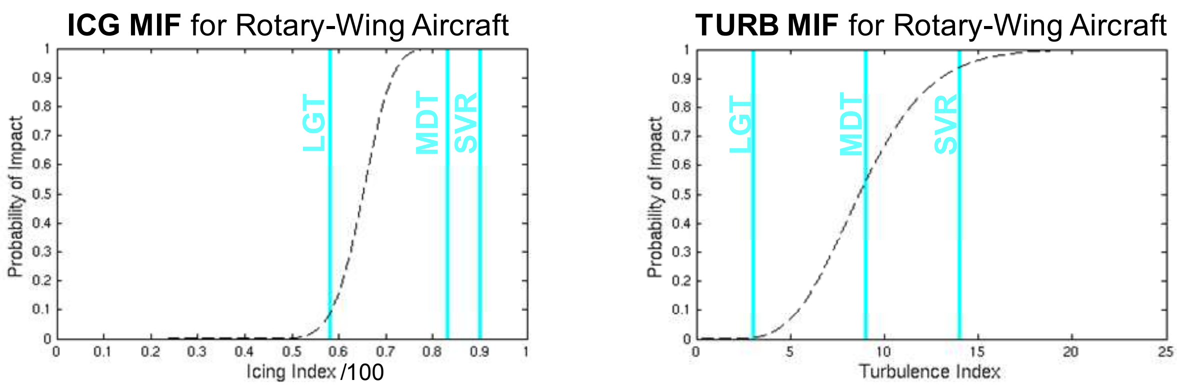


Figure 12. Default MIF for ICG and TURBC. Standard thresholds for light (LGT), moderate (MDT), and severe (SVR) intensities are shown for reference.

These MIFs are mainly intended to demonstrate *WIPCast* processing and produce a reasonable result. The TURBC MIF is defined by a gamma CDF with $\mu = 9.0$ and $\sigma = 3.0$, which equates to gamma parameters of $\alpha = 9.0$ and $\beta = 1.0$ using the method of moments. The ICG MIF is defined by a beta CDF with $\mu = 0.65$ and $\sigma = 0.05$, which equates to beta parameters of $p = 58.5$ and $q = 31.5$ also by the method of moments.

A key question being addressed as part of *WIPCast* development is how to appropriately define a meaningful MIF for each weather vulnerability. A reasonable approach is to assume a basic shape of the PDF (Gaussian or gamma) and then define its first two moments (mean and standard deviation) by assigning assumed percentile values to the T-IWEDA marginal and critical thresholds, thus grounding the MIF on known information. This process was shown graphically in Figure 3. b where the marginal threshold was placed at the 10^{th} percentile and the critical threshold was placed at the 50^{th} percentile.

3.2.6 Decision Support / User Interface

Our Phase I effort has incorporated preliminary prototyping of *WIPCast* displays that intuitively and clearly present anticipated adverse weather impacts and risks at a grid point, along a flight leg, and along an entire flight path. The client interface also provides a mechanism for the user to characterize their risk tolerance, which then helps drive the information presented in the resulting decision support displays.

The Phase I prototype user interface is shown in Figure 13. .

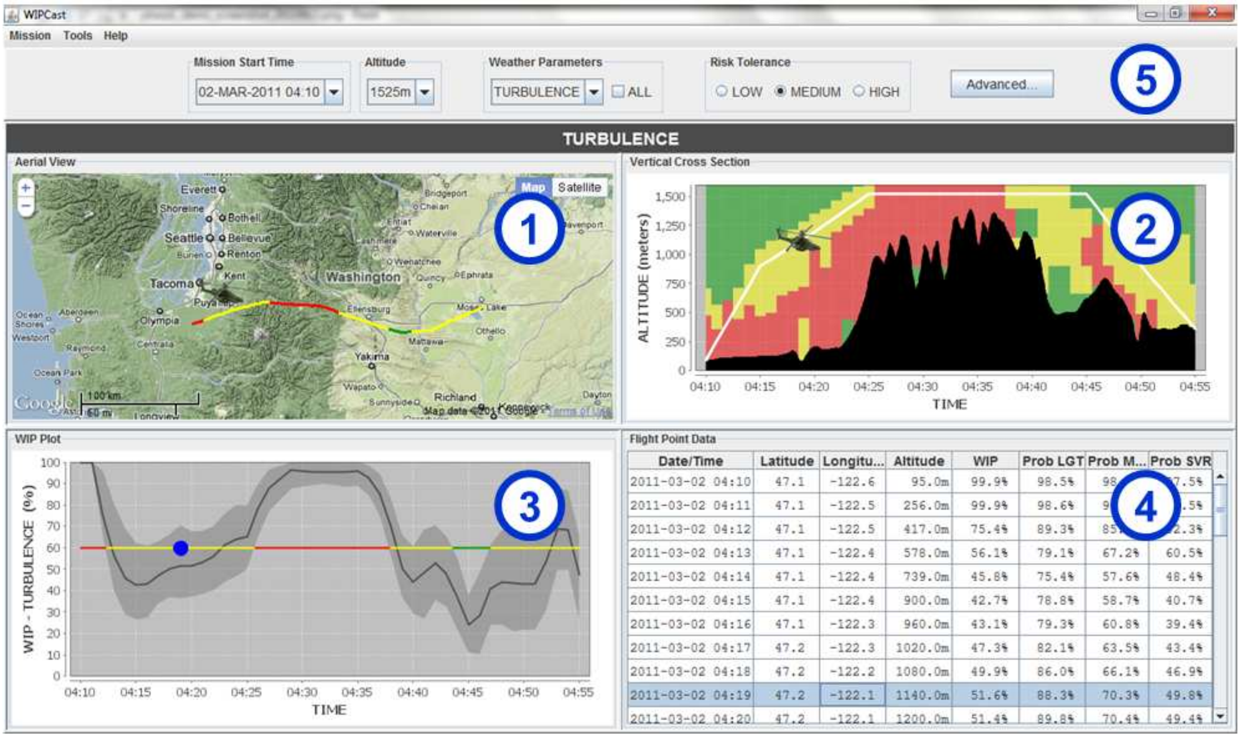


Figure 13. WIPCast Phase I Prototype User Interface

1 ***Geospatial (Aerial) View.*** Provides color-coded display of operational risk along entire flight path. Fully integrated with Google Maps.

2 ***Vertical Cross Section (Altitude View).*** Provides color-coded grid cell display of operational risk at all altitudes along entire flight path, as well as displaying terrain along entire flight path.

3 ***WIP Graph.*** Plots calculated WIP spread (variance) along entire flight path. Graphically depicts relationship between WIP values and user-selected risk tolerance. Helps answer the question of "how red is red".

4 ***Flight Point Data Display.*** Provides detailed location (lat/lon/alt) and weather data for each defined step along the flight route.

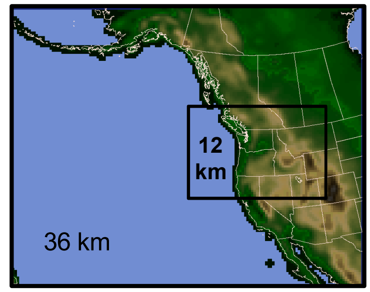
5 ***User Controls.*** Allow for interactive selection of mission start time, flight route (path / altitude), weather parameters of interest, and operational risk tolerance. Advanced controls are also available for adjusting impact thresholds (as input to MIF).

All views are fully integrated and synchronized, updating dynamically in response to changes in user selections (e.g., risk tolerance, mission start time, etc.).

3.2.7 Mesoscale Ensemble for Phase I

While the *WIPCast* solution is being carefully architected to avoid reliance on a particular ensemble system, the Phase 1 proof-of-concept prototype makes use of the high-resolution (12-km) University of Washington Mesoscale Ensemble (UWME), a mesoscale probabilistic testbed that has run in real-time for nearly a decade. Although UWME does have known limitations (such as being limited to nine members and not addressing model uncertainty), it is still a useful basis for the Phase 1 prototype given its high resolution. Moreover, UMWE has provided data for Phase I covering the Pacific Northwest, a region of substantial meteorological contrasts, with a wide-range of surface conditions and climate types that enabled us to better explore the sensitivities of *WIP* during Phase I. Our open *WIPCast* architecture retains the option in Phase II of ingesting ensemble data from other sources, such as the Air Force Weather Ensemble Prediction System (AFWEPS) run by the Air Force Weather Agency, the primary supplier of weather information to the US Army.

UWME is an ensemble of nine short-range numerical weather prediction (NWP) model forecasts run operationally at the UW Department of Atmospheric Sciences. Forecasts are run twice per day (at the 00 and 12 UTC cycles) for the 0-72 hour forecast period to produce skillful probabilistic mesoscale meteorological forecasts for the Pacific Northwest United States, using the Weather Research and Forecasting (WRF) model.



The forecasts feature an outer grid (151x127) of 36 km horizontal grid spacing that covers much of western North America and the northeastern Pacific and a nested grid (163x124) of 12 km grid spacing that covers the entire Pacific Northwest U.S., from central California to central British Columbia, and from 600 miles off the west coast to central Montana, as shown here. The model utilizes 37 vertical eta levels, and is non-hydrostatic, which limits pressure gradient force errors in complex terrain. An upper-radiative boundary condition is used to allow gravity waves to radiate through the model top without being reflected. The sub-grid scale parameterizations include:

- Yonsei University planetary boundary layer scheme
- Thompson explicit moisture scheme including mixed phase processes
- Kain-Fritsch cumulus parameterization
- Simple shortwave and RRTM longwave radiation schemes
- Noah land surface model

UWME employs a multi-analysis strategy to provide the forcing of its nine members. Initial conditions (ICs) and lateral boundary conditions (LBCs) for the WRF mesoscale ensemble are interpolated from separate synoptic-scale model analysis and forecast fields obtained from several operational weather prediction centers worldwide. In addition to providing the LBC updates, nudging by the synoptic-scale forecasts is also applied to the 36 km (outermost) model domain to ensure consistency with the prescribed synoptic forcing. The nine synoptic-scale ICs and LBCs includes:

1. The National Center for Environmental Prediction (NCEP) Global Forecast System (GFS)
2. The Canadian Meteorological Centre (CMC) Global Environmental Multiscale (GEM) model
3. The National Center for Environmental Prediction (NCEP) Eta model
4. The Bureau of Meteorology (Australia) Global AnalysiS and Prediction (GASP) model
5. The Japan Meteorological Agency (JMA) Global Spectral Model (GSM)
6. The Fleet Numerical Meteorology and Oceanography Center (FNMOC) Navy Operational Global Analysis and Prediciton System (NOGAPS)
7. The Central Weather Bureau (CWB) (Taiwan) Global Forecast System
8. The United Kingdom Met Office (UKMO) Unified Model (UM)
9. Centroid (CENT) – average of the above 8 analyses/forecasts

Forecasts are computed on a 16-node (116 Intel Xeon 2.3GHz processors) Linux PC Beowulf cluster. Each 36/12 km 72-hour forecast finishes in approximately 2 hours. Each global/synoptic-scale IC becomes available at a different time after 0000/1200 UTC. Thus, some UWME ensemble members finish before others. UWME processing is completed by 1200/0000 UTC (0400/1600 PST) provided all initializations are in on time.

3.2.8 DATA COLLECTION AND INGEST

The University of Washington Mesoscale Ensemble (UWME) was used as the source of both the ensemble forecast data and the ground truth (i.e., verifying observations) for *WIPCast* Phase I. The raw ensemble model output data was vertically interpolated from eta levels to fixed pressure levels, with 25 mb vertical resolution from 1000 to 700 hPa, and 50 hPa resolution from 700 to 100 hPa. The archive of ensemble data for the *WIPCast* Phase I includes surface fields and variables describing the basic state of the atmosphere at pressure levels, including wind, temperature, humidity, and geopotential height. The archive extends from December 21, 2010 through present.

While it is best to use a high quality independent source for ground truth data (used for calibration and verification), *WIPCast* Phase I used short-term forecasts from UWME's Centroid member as ground truth in order to stay within budget constraints. For example, a UWME forecast with a lead time of 28h was verified against the 4-h Centroid forecast from the most recent forecast initialization. This approach has some validity since updated, short-term forecasts should be closer to reality than forecasts with greater lead times, but it of course suffers from the fact that short-term forecasts can still carry significant errors. Also, the model's systematic errors are shared between both the forecast and ground truth so it is difficult to properly calibrate the forecasts. Note that these limitations do not impact the fundamental soundness of *WIPCast*'s processing algorithms which are designed independent of the quality of the ground truth. A more robust ground truth can easily be plugged into the framework.

3.2.9 TECHNICAL IMPLEMENTATION

The *WIPCast* prototype has been carefully designed to serve as a firm foundation for development of the operational system in Phase II.

The front-end developed by Impact Computing has been written entirely in Java and is therefore cross-platform. Impact believes in employing open source software where feasible and appropriate to avoid reinventing the wheel. Accordingly, two open source projects – the NetCDF-Java library and the JFreeChart library – were leveraged in development of the Phase I prototype. (As part of the Phase I development effort, Impact identified bugs in the JFreeChart package for which Impact submitted fixes to the JFreeChart open source project on SourceForge.)

The back-end software developed by University of Washington has been written in Fortran and Perl, running on a Linux cluster with 16 nodes and a 40 TB archive disk array.

3.3 PHASE I TASKS

The specific tasks performed during Phase I included the following:

Task 1. Ingest and processing of mesoscale ensemble model output

- Captured high-fidelity archive of ensemble forecast data
- Developed code to analyze data source performance characteristics, for ensemble calibration and ambiguity calculation training purposes

Task 2. WIP calculation

- Produced forecast for aircraft icing using a derived "icing index" variable and for aircraft turbulence using a derived "turbulence index" variable
- Performed calibration of ensemble forecast data for both Gaussian and non-Gaussian variables
- Performed flight route interpolation
- Implemented "stretching" of ensemble distributions, separately at each vertical data level

- Generated sample probability forecasts
- Designed sample Mission Impact Function (MIF) for potential effect of both icing and turbulence on Army aviation
- Implemented technique for estimation of ambiguity in probability forecast and in WIP calculation
- Calculated WIP for individual variables (turbulence, icing) as well as “total WIP” (i.e., joint probability of weather impact from both turbulence and icing)

Task 3. Software architecture and design

- Architected and designed back-end server for generation of WIP data
- Architected and designed front-end client for retrieval, processing and display of WIP data in mission context
- Designed client/server interface
- Defined data file specification

Task 4. User interface design

- Designed *WIPCast* prototype user interface to support the mission planner use case
- Designed integrated graphical and tabular displays of WIP data
- Designed intuitive integrated displays for graphical representation of anticipated adverse weather impacts and risks at a grid point, along a flight leg, and along an entire flight path
- Designed controls for setting user preferences and mission parameters

Task 5. Proof-of-concept prototyping

- Developed a preliminary proof-of-concept working prototype, capable of:
  - Generating *WIPCast* data from probabilistic forecasts
  - Using *WIPCast* data to drive a decision support tool for the mission planner
- Prepared test dataset for use with prototype
- Extracted interpolated data along sample flight route

Task 6. Project management

3.4 PHASE I OPEN ISSUES

The following open issues have emerged from the Phase I effort as areas requiring for research in Phase II (see Section 5.2.1):

- ***Turbulence Forecast.*** The *WIPCast* approach to producing turbulence forecasts employed in Phase I produced good results at upper levels (>10,000ft) but often considerably over-forecasted turbulence in the low levels. An alternate approach that is expected to effectively address this issue has been identified and needs to be implemented in Phase II.
- ***Representation of Ground Truth.*** Current use in *WIPCast* of a short-range model forecast for representation of ground truth has serious limitations in realism and can present a biased view. An independent, unbiased (or at least only slightly biased) ground truth with a sufficiently fine spatio-temporal resolution covering the entire forecast domain is needed.

- ***Ambiguity Calculation.*** *WIPCast* needs to improve both accuracy and efficiency of ambiguity calculation. An alternate approach that is expected to effectively address this issue has been identified and needs to be implemented in Phase II.
- ***Calibration.*** The validity and skill of the calibration needs to be analyzed. While the shift-and-stretch method has been shown to be effective in general (Eckel and Allen 2011), it is unclear how it works for the derived variables used in *WIPCast*, particularly with respect to the adaptation to non-Gaussian variables.
- ***Weighting Schemes.*** It is important for *WIPCast* to accommodate the ability to specify relative weights of "importance" for individual weather sensitivities, especially since relative importance of individual weather events is likely to vary based on type of mission and other operational constraints. A particularly interesting outcome of our Phase I research was the realization that user input in defining impact thresholds (as input to the MIF) serves as an implicitly effective means for relative weighting of weather sensitivities, which is the approach we therefore plan to pursue in Phase II.
- ***Decision Support.*** The Phase II effort will extend the innovative user interface paradigms developed in Phase I to further enhance the decision support capabilities provided by the *WIPCast* client display.
- ***3D Visualization.*** One of the visualization approaches explored in Phase I was that of a 3D cube of the entire region encompassing the mission. Although such a display would appear conceptually to be beneficial, we found it particularly challenging in Phase I to devise a 3D visualization and interface paradigm that was sufficiently intuitive and useful so as to add value beyond our multiple integrated 2D displays. We nevertheless remain hopeful that some form of 3D display will in fact prove useful as part of the *WIPCast* user interface and therefore intend to explore this area further in Phase II.

3.5 PHASE I TEAM

The Phase I effort was spearheaded by **Impact Computing**, with Hyam Singer serving as Principal Investigator, supported primarily by Mr. Zev Hochberg. Extensive support was provided by the **University of Washington's Department of Atmospheric Sciences**, under the leadership of Dr. Cliff Mass, supported by Mr. Rick Steed and Mr. Jeff Baars.

The seminal work of Dr. Tony Eckel in the domain of probabilistic forecasting and its practical application has served as the foundation for much of our WIPCast solution. Our team has therefore benefited greatly from the close collaboration and guidance of Dr. Eckel, who has provided invaluable input and feedback throughout our Phase I effort.

Brief bios of the active Phase I team members are provided below.

MR. HYAM SINGER, IMPACT COMPUTING CORP., PRINCIPAL INVESTIGATOR

Mr. Singer, Impact's President and founder, has over 30 years of experience in the software industry and has successfully managed the development of a wide array of innovative, operationally-relevant solutions for government and commercial customers. Mr. Singer's broad range of technical expertise spans the areas of visualization, user interface design, geographic information systems (GIS), image exploitation, modeling and simulation (M&S), data mining, and complex systems architecture. Mr. Singer possesses a proven track record as a creative and results-oriented professional, highly skilled at spearheading innovative solution development, from conceptualization through deployment in the field. Over the course of his career, Mr. Singer has served as Principal Investigator for a number of highly successful Phase 1 and Phase 2 SBIR initiatives, including a prior SBIR effort for ARL. Mr. Singer possesses a B.S. in Computer Science from the University of Maryland and an active Top Secret security clearance.

MR. ZEV HOCHBERG, IMPACT COMPUTING CORP., SENIOR SYSTEMS ANALYST

Mr. Hochberg is a technical innovator with over 25 years of progressive technical and leadership experience in software engineering. During his tenure at Impact, Mr. Hochberg has served as lead developer on a number of SBIR initiatives in addition to *WIPCast*, including a wiki-based collaborative analysis and information sharing framework for complex datasets, and a spatio-temporal data model that captures and represents the dynamics inherent in spatio-temporal datasets over both space and time. Mr. Hochberg has also served as a key member of Impact's eXtensible Behavioral Modeling (XBM) development team, responsible for developing major portions of this flexible open source framework for integrating a hybrid combination of modeling and simulation tools. Mr. Hochberg possesses an M.S. in Computer Science from New York University's Courant Institute as well as an M.A. in Physics from Yeshiva University's Belfer Graduate School of Science.

DR. CLIFFORD MASS, UNIVERSITY OF WASHINGTON, MESOSCALE MODELING EXPERT

Dr. Mass, Professor of Atmospheric Sciences at the University of Washington (UW) and Lead Scientist of the Northwest Modeling Consortium, possesses over 20 years of experience in mesoscale prediction, and has run a real-time mesoscale ensemble prediction effort for over 10 years. With colleagues in the UW statistics and psychology departments, he has built an end-to-end mesoscale prediction system including sophisticated post-processing and displays. He was a member of the NRC Committee that wrote the report on the need for uncertainty information in weather prediction. Dr. Mass possesses a Ph.D. in Atmospheric Sciences from the University of Washington. Among his numerous refereed publications, a sampling of a few most relevant to this initiative includes:

Kleiber, W., A. E. Raftery, J. Baars, T. Gneiting, C. F. Mass, and E. Grimit, 2010: Locally Calibrated Probabilistic Temperature Forecasting Using Geostatistical Model Averaging and Local Bayesian Model Averaging. Submitted to Mon. Wea. Rev.

Mass, C., S. Joslyn, J. Pyle, P. Tewson, T. Gneiting, A. Raftery, J. Baars, J. M. Sloughter, D. Jones and C. Fraley, 2009: PROBCAST: A Web-Based Portal to Mesoscale Probabilistic Forecasts. Bulletin of the American Meteorological Society, 90, 1009–1014

Grimit, E. P., and C. F. Mass, 2007: Measuring the Ensemble Spread–Error Relationship with a Probabilistic Approach: Stochastic Ensemble Results. Monthly Weather Review,135, 203–221

Eckel, F. A. and C. F. Mass, 2005: Effective mesoscale, short-range ensemble forecasting. Weather and Forecasting, 20, 3238-350

MR. RICK STEED, UNIVERSITY OF WASHINGTON, RESEARCH METEOROLOGIST

Mr. Steed possesses 15 years of experience in mesoscale numerical weather prediction. For the past 10 years he has constructed and maintained the UW mesoscale ensemble prediction system. During this time he participated in numerous projects which developed applications of ensemble weather forecast data to probabilistic decision-making tools. Mr. Steed possesses an M.S. in Atmospheric Sciences from the University of Washington.

MR. JEFF BAARS, UNIVERSITY OF WASHINGTON, RESEARCH METEOROLOGIST

Mr, Baars, Research Meteorologist at the University of Washington, possesses over 15 years of experience, including the last nine years working with Dr. Mass on numerical weather prediction, verification, and post-processing. Mr. Baars played a key role in the creation of Probcast, a pioneering probabilistic weather prediction system, as well as enhancing and implementing Probcast algorithms on the AFWA Joint Ensemble Forecast System. Mr. Baars possesses an M.S. in Atmospheric Sciences from the Ohio State University.

4 PHASE I OPTION

Our intent is to use the Phase I Option Period to refine the Phase I results so as to further set the foundation for Phase II. Likely focus areas for the Phase I Option Period activities will include:

- Multivariate calculations and analysis
- Additional visualization techniques and enhancements to the user interface
- Availability and use of AFWA (Armed Forces Weather Agency) data
- *WIPCast* validation and verification (V&V) techniques, including source of "ground truth" data
- Web (i.e., browser-based) client
- GoogleEarth integration

Final selection of Phase I Option Period activities will be based on customer input and prioritization.

5 PHASE II PLAN

We have carefully designed and implemented the Phase I working prototype of our *WIPCast* solution to serve as a firm technical foundation for a subsequent Phase II effort. In addition, our Phase I research – along with feedback and guidance from the customer – has yielded insights highly relevant to design and development of the Phase II *WIPCast* solution.

The ultimate objective of the proposed follow-on Phase II effort is to develop an operational version of the *WIPCast* system, building upon our successful design and prototyping activities conducted during Phase I. A functional overview of our vision for the Phase II system is depicted in Figure 14. .

.

Figure 14. WIPCast Phase II Functional Overview

5.1 PHASE II PLAN OVERVIEW

The following is a high-level overview of our Phase II plan:

- **Goal:** Fully operational "version 1.0" of the WIPCast system
- **Period of performance:** 24 months
- **Deliverables:** WIPCast 1.0 software (open source), Technical Documentation, Monthly Status Reports, Final Report
- **Schedule:** Iterative development, with interim customer meetings and deliverables every 6 months
- **Team:** All members of Phase 1 team, expanded to include ZedX (weather data processing and modeling) and Marcus Weather (commercialization)

5.2 PHASE II TECHNICAL OBJECTIVES

Specific Phase II technical objectives, as detailed in the subsections that follow, consist of addressing Phase I open issues, addressing key requirements of the operational *WIPCast* system, and laying the groundwork for *WIPCast* productization.

5.2.1 ADDRESSING PHASE I OPEN ISSUES

A key goal of the Phase II effort will be to further research and effectively address the following open issues (previously detailed in Section 3.4) that emerged from the Phase I effort:

- ***Turbulence Forecast.*** See Section 5.3.9.
- ***Representation of Ground Truth.*** See Section 5.3.8.
- ***Ambiguity Calculation.*** See Section 5.3.4.
- ***Calibration.*** See Section 5.3.5.
- ***Weighting Schemes.*** See Section 5.3.7.
- ***Decision Support.*** See Section 5.3.10.
- ***3D Visualization.*** See Section 5.3.10.

5.2.2 Key Operational Requirements

In addition to addressing the open technical issues enumerated in Section 5.2.1, there are a number of key operational requirements that must be satisfied to achieve a version of the *WIPCast* system that is sufficiently marketable and deployable, including:

- ***Support for additional weather variables.*** While the focus on turbulence and icing was entirely adequate for our Phase I proof-of-concept prototyping efforts, the Phase II system will need to incorporate support for additional weather variables of interest. See Section 5.3.1.
- ***Dynamic WIP Calculation.*** *WIPCast* needs to build the capability to recompute output on-the-fly based on dynamic user or external inputs regarding flight route (path, time, altitude), weather sensitivity impact thresholds (for defining the MIF), and risk tolerance. See Section 5.3.2.
- ***Multivariate WIP Calculation.*** The chance of a negative impact depends on multiple elements, such as wind speed and direction, visibility, turbulence, and icing. In addition, it depends on these quantities at multiple points in time and space across the mission. Thus, in practice, the total WIP depends on a large number of inter-related weather phenomena simultaneously and therefore necessitates the ability to address the multivariate WIP challenge. See Section 5.3.3.
- ***Integration with External Systems and Data Sources.*** Formal definition of the *WIPCast* Server APIs (Application Programming Interfaces) in Phase II will facilitate integration with other systems such as MyWIDA, T-IWEDA, JMPS, and AWRT. These systems will thereby be able to exploit outputs of the *WIPCast* system, as well as providing mission configuration and other inputs to *WIPCast*. In addition, the *WIPCast* front-end geospatial display, which was integrated with GoogleMaps in Phase I, will also be integrated with GoogleEarth in Phase II. See Section 5.3.11.
- ***Browser-based Client.*** The Phase I prototype provided a desktop client. Also providing a web browser-based client as an alternative provides important operational benefits, such as ease of deployment and accessibility. See Section 5.3.10.

5.2.3 Productization

With an eye toward productization and commercialization, the Phase II effort is also to yield:

- Comprehensive user and technical documentation
- Suite of automated functional and regression test scripts
- Verification and validation (V&V) plan

Further detail is provided in Sections 5.3.12 and 5.3.13.

5.3 Phase II Tasks and Research Areas

The specific tasks to be performed and research areas to be addressed in Phase II are detailed in the subsections that follow. Each task directly supports one or more of the technical objectives listed in Section 5.2.

5.3.1 Support for Additional Weather Variables

While the focus on turbulence and icing was adequate for our Phase I proof-of-concept prototyping efforts, the Phase II system will need to incorporate support for additional weather variables of interest.

In support of the Army aviation use case, a Phase II test case will most likely be selected that focuses on a specific aircraft (e.g., Apache) and extends WIPCast capabilities beyond icing and turbulence to also supporting at least 3 additional weather parameters of interest (such as cloud ceiling, visibility, winds, precipitation, etc.).

5.3.2 DYNAMIC WIP CALCULATION

WIPCast needs to build the capability to recompute output on-the-fly based on dynamic user or external inputs regarding flight route (path, time, altitude), weather sensitivity impact thresholds (for defining the MIF), and risk tolerance. For a change only in risk tolerance, *WIPCast* will just update the R-Y-G decision recommendations along the route. For a change in impact thresholds, *WIPCast* will update the MIF so as to recompute WIP and its 90% CI (using RCR) along the flight route. For a change in flight route, *WIPCast* will have to recompute everything beginning with interpolation of the original ensemble forecast data along the route.

5.3.3 MULTIVARIATE WIP CALCULATION

Meeting the challenge of accurately calculating a total *WIP* that accounts for concurrent risks from multiple weather sensitivities has been deferred to Phase II. For Phase II, we plan to first develop an accurate method for calculating total (i.e., multivariate) *WIP* starting with the relatively simply two-event (i.e., icing and turbulence) problem. After building a foundation with the two-event problem, we will expand to handling more variables (slant range visibility, cloud ceiling, wind speed, etc.) simultaneously.

In the Phase I prototype, WIP was generated separately for two weather sensitivities (turbulence and icing) and a total WIP was produced that accounts for both risks using a simplification. To understand the simplification, as well as the larger challenge, we present the following basic explanation.

First, consider just the two event (i.e., icing and turbulence) problem. We define **A** as being an impact from TURBC, so **Pr(A)** is *WIP* from TURBC, and **B** as being an impact from ICG, so **Pr(B)** is *WIP* from ICG. The total *WIP* is then the chance of either impact (**A** or **B**) occurring:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

Calculating the last term on the right – $\Pr(A \cap B)$ – is where the challenge lies. It is known as the joint probability and is best understood with a Venn diagram:

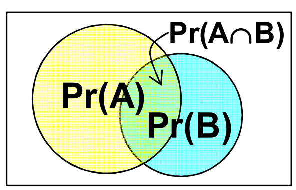


Figure 15. Venn diagram example of two-event situation with Pr(A) = 0.4 and Pr(B) = 0.3

The joint probability is the overlap between the two events' probabilities, within the space of all possible outcomes (i.e., the complete box). Total WIP is the fraction of the box that is shaded any color. So as to avoid counting the overlap area (i.e., the green area) twice, the equation above subtracts the overlap from the total of the two event probabilities. However, calculating the actual value of the joint probability is anything but simple:

$$\Pr(A \cap B) = \Pr(A)*\Pr(B|A)$$

or equivalently:

$$\Pr(A \cap B) = \Pr(B)*\Pr(A|B)$$

The conditional probability, PrA|B), is normally not known and difficult to estimate when the two events have a variable degree of correlation (i.e., are not independent). This is the case for TURBC and ICG (the two variables used in the Phase I prototype) since they often go hand-in-hand (e.g., a stronger storm has both more intense TURBC and ICG), but can also correlate weakly at other times. Note that the correlation of interest here is on occurrence of an event (i.e., binary Y/N for impact from TURBC or ICG) and not correlation between two continuous random variables.

Substituting in the expression for joint probability to the total *WIP* equation, we simplify to:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) * [1 - \Pr(A|B)]$$

This provides a logical perspective for the challenge. The total *WIP* is found by starting with the larger of the two *WIP*s and adding on part of the smaller *WIP*, so the minimum total *WIP* is Pr(A). The fractional part of Pr(B) to add on is related to the correlation between A and B, as seen in Figure 16. .

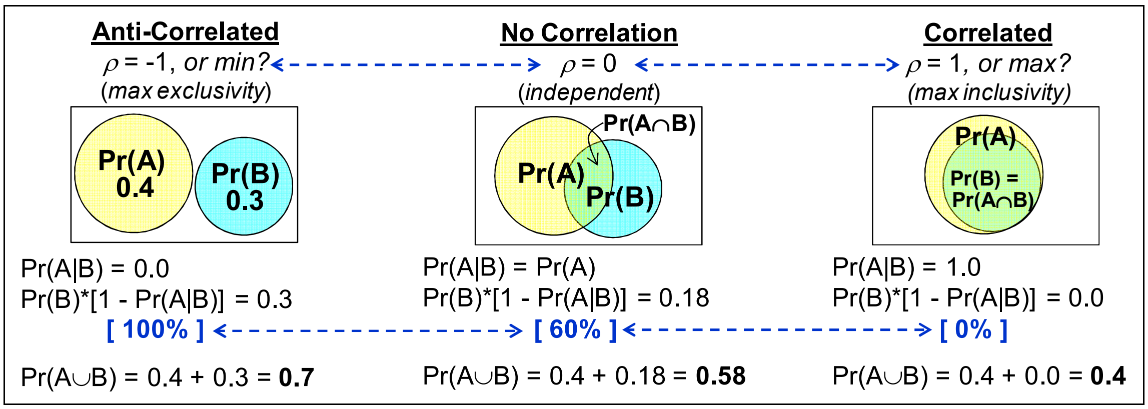


Figure 16. Depiction of total WIP calculation, Pr(A∪B), over the continuum of possible amounts of correlation (indicated by the correlation coefficient, ρ) between A and B. The blue % values are the fraction of Pr(B) that must be added to Pr(A) to arrive at the total WIP.

In the anti-correlated case, when both probabilities are small enough, they can be mutually exclusive and have ρ = -1 as shown here. So for total *WIP*, all of Pr(B) gets added on to Pr(A). Once the sum of the two probabilities is > 1.0, there will be some overlap, both Pr(A|B) and ρ will be at some minimum value, and Pr(A∪B) = 1.0.

On the opposite extreme, when A and B are highly correlated, they can be perfectly correlated only if they are equal, otherwise ρ will be at some maximum value. However Pr(A|B) always equals 1.0 since we chose the perspective that Pr(A) > Pr(B). So there is nothing to add onto to Pr(A) from Pr(B) and total *WIP* is simply Pr(A).

In the middle case, A and B occur independently but can still occur at the same time. The joint probability simplifies to Pr(A)*Pr(B) and the conditional probabilities are equal to the probability of the second event. This is the case we assumed in calculating total *WIP* for Phase I.

To emphasize the importance of risk analysis with total WIP, consider a decision by a user with a risk tolerance of 45% who is given the risks depicted in Figure 16. . Risk analysis performed separately on each of the two possible events would result in an overall 'Go' decision since risk is acceptable from both TURBC and ICG individually. However, research shows that ρ = 0.2 at these levels of risk, so total WIP = 0.54 and the appropriate decision is 'No-Go' since there is too much risk.

For Phase II, we plan to further explore the continuum depicted in Figure 16. to develop an accurate method of calculating total *WIP*, starting with the relatively simply two-event problem. The basic approach is to find either the correlation between the two events and relate it to the joint probability, or find the conditional probability directly. Note, however, that this involves exploring the events of TURBC or ICG (at or above a selected intensity level) rather than the true end goal of exploring the events of an *impact* from TURBC or ICG. Since we are unable to track the event of actual impacts from TURBC or ICG, we will track occurrence of TURBC and ICG, assume an impact, and transfer the resulting relationships to apply to *WIP*.

There are many possible dependencies to explore for possible ways to condition the variability in the correlation between the two events. A prominent one is simply that the correlation depends on the amount of probability itself. Large probability values naturally must carry high correlation, but low probability values may go either way depending upon the variables in question. Another dependency to examine is the weather regime. The correlation between the events may change dramatically over different seasons, atmospheric stability, topography, etc. Lastly, the relative intensity of the event may play a role since light-moderate conditions could conceivably vary independently while severe conditions go more hand-in-hand.

After building a foundation with the two-event problem, we will expand to handling more variables (slant range visibility, cloud ceiling, wind speed, etc.) simultaneously. While we recognize that the complexity increases substantially with more variables (since it involves 3-way, 4-way, etc. correlations or conditional probabilities), the risk-analysis decision still ultimately comes down to comparing the user's risk tolerance to the best possible approximated estimate of the total *WIP*.

5.3.4 IMPROVED AMBIGUITY CALCULATION

WIPCast needs to improve both accuracy and efficiency of ambiguity calculation. Toward that end, in Phase II:

- The RCR code will be completed to include the random calibration component.
- As a huge boost to efficiency, the RCR code will be adapted to run *only* along the flight route (which will require interpolation of all ensemble member data along the route).
- The higher-performance CES method for ambiguity calculation will then be implemented to generate 90% confidence intervals (CI) for forecast probability and WIP over the complete domain.

While there will be some disagreement between the RCR and CES results, *WIPCast* will employ CES to convey a general, overall view of the ambiguity while RCR will be used to present a more refined view where it is needed (e.g., along the flight route).

5.3.5 CALIBRATION

The validity and skill of the calibration needs to be analyzed. While the shift-and-stretch method has been shown to be effective in general (Eckel and Allen 2011), it is unclear how it works for the derived variables used in *WIPCast*, particularly with respect to the adaptation to non-Gaussian variables.

A thorough verification analysis is to be performed in Phase II, the results of which will likely reveal multiple alternatives for further improving the calibration, such as conditioning the 2^{nd} moment correction on the ensemble spread (Eckel and Allen 2011).

5.3.6 MIF SPECIFICATION AND WEIGHTING OF WEATHER SENSITIVITIES

One approach for defining a MIF is to allow for input from the users. The idea is that the user has knowledge of the factors that can affect the likelihood of impact (e.g., flight crew experience, aircraft load, aircraft age, etc.) for a given observed weather condition, and should thus adjust the MIF accordingly. For example, a very experienced pilot may be able to handle MDT TURBC much better than a novice pilot and thus greatly increase the odds of mission success when MDT TURBC occurs. The user would then be justified in adjusting the corresponding TURBC MIF threshold to be higher.

Varying the impact thresholds, and thereby the MIF, can greatly change the final value of *WIP*, perhaps more accurately reflecting the overall risk to the mission. Rather than dealing with a complex CDF, *WIPCast* will allow a user to specify the weather level at which there is a minor chance of impact and a major chance of impact (Figure 17.), based on their knowledge of factors such as the skill and fatigue of the team, the age of the equipment, and so on. There are then assumed to be the 10th and 90th percentiles of the MIF which are calculated behinds the scenes. The MIF thereby represents the uncertainty in an impact from actual weather due to all remaining unclear factors.

Further research with regard to how to appropriately define a MIF for each weather sensitivity is also planned for Phase II.

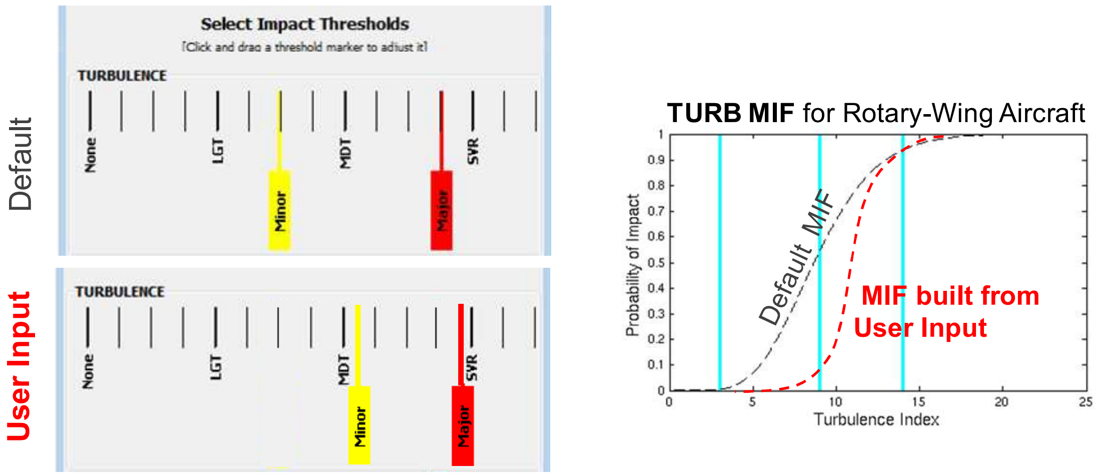


Figure 17. User interface for selecting impact thresholds. Top left image shows default settings for turbulence impact thresholds. Lower right image shows sample adjustment by user, with the resulting MIF shown in the graph on the right.

5.3.7 Weighting Schemes

For any given mission, a variety of factors may cause certain events to be more important or significant to mission success than others. Accordingly, it is important for *WIPCast* to accommodate the ability to specify relative weights of “importance” for events. However, as discussed in our Phase I proposal, there are significant obstacles to defining a weighting scheme that has anywhere near universal applicability or that will be equally agreeable to a large population of users. In particular:

- Relative importance of individual events is likely to vary based on type of mission and other operational constraints
- Given the interdependence of multiple events, a simple relative weighting scheme may not be adequate to address the complexities of event prioritization
- Individual users may not agree on the appropriate relative weights to apply to relevant events, even in the same operational scenario

A particularly interesting outcome of our Phase I research was the realization that user input in defining impact thresholds (as input to the MIF) serves as an implicitly effective means for relative weighting of weather sensitivities. (e.g., risk from turbulence, lightening, low visibility, etc.). The higher the sensitivity to a particular weather condition, the lower the corresponding impact threshold will be set by the user (since the threshold determines the point beyond which a negative impact is anticipated), and vice versa. For example, if a user is more concerned about the potential impact from icing than the impact from turbulence, he/she would then adjust the thresholds accordingly (i.e., lowering the impact threshold for icing to indicate greater sensitivity and/or raising the impact threshold for turbulence to indicate lower sensitivity), thereby implicitly giving icing heavier weighting than turbulence in total WIP calculation. These implied relative weights will then carry over into the calculation of total *WIP*, which will thereby reflect a weighted, total chance of impact.

5.3.8 WEATHER OBSERVATION DATA ("GROUND TRUTH")

Current use in *WIPCast* of a short range model forecast for representation of ground truth has serious limitations in realism and can present a biased view. An independent, unbiased (or at least only slightly biased) ground truth with a sufficiently fine spatio-temporal resolution covering the entire forecast domain is needed. Assuming a CONUS forecast domain, the target for surface variables is the Real Time Mesoscale Analysis (RTMA), which is currently produced hourly on a 5km grid and will soon be upgraded to a 2.5km grid. RTMA data is available for free from the NWS. For non-surface data, the target is the Rapid Update Cycle (RUC) analysis, which is produced hourly on a 13km grid covering CONUS. RUC13 data is available from NOAA.

5.3.9 IMPROVED TURBULENCE FORECAST

The turbulence index (*TI*) algorithm employed in Phase I followed the method presented by Ellrod and Knapp (1992), which considers the wind shear, deformation and convergence in the forecasted, three-dimensional model wind field. Given the UWME model resolution of 12km, we followed AFWA's empirically based thresholds for mesoscale models of LGT ($3 < TI \leq 9$), MDT ($9 < TI \leq 14$), and SVR ($14 < TI$). This produced good results at upper levels (>10,000ft) but often considerably over-forecasted turbulence in the low levels.

This issue can be alleviated by implementing the Panofsky Index (PI) to represent turbulence in the lower atmosphere (below about 5,000ft AGL) (Boyle 1990, Passner 2000, Brooks and Oder 2004). A challenge will be how to smoothly blend from PI to TI to avoid any discontinuities in the final product. One possibility is to leverage how AFWA deals with that challenge since they use these indices in operations.

5.3.10 DECISION SUPPORT AND USER INTERFACE

The most powerful and well architected systems are often rendered essentially useless by an ill-designed user interface. As Donald Norman states in his insightful book The Design of Everyday Things, "There is a big difference between the expertise required to be a designer and that required to be a user." Norman goes on to state that, whereas designers become expert with the device they are designing, users become expert at the task they are trying to perform with the device. These two perspectives are unfortunately often worlds apart. At Impact Computing, our approach to user interface design is therefore highly "user-centric", relying heavily on an iterative cycle of rapid prototyping and user feedback to ensure a solution that is operationally relevant, intuitive and easy-to-use.

Accordingly, the Phase I prototype's *WIPCast* display intuitively and clearly presents anticipated adverse weather impacts and risks at a grid point, along a flight leg, and along an entire flight path. The interface also provides a mechanism for the user to characterize their risk tolerance, which then helps drive the information presented in the resulting decision support displays.

Building on this foundation, Phase II research and development tasks relating to the *WIPCast* decision support front-end will include:

- Creation of a **mission specification "wizard"** enabling users to easily input mission configuration parameters
- **Additional 2D displays**, such as:
    - A graphical representation of predicted weather forecast values as compared with corresponding parameter threshold values (this display was suggested by the customer at our final Phase I meeting and would provide an useful added dimension of information complementing the existing WIP graph display developed in Phase I)

  - A dashboard (i.e., “stop light”) display summarizing the anticipated impacts and risk levels for each of the weather parameters being evaluated

- **3D cube** of the entire region encompassing the mission[5]
- **Additional spatio-temporal controls** to allow for further manipulation of the *WIPCast* displays over both space and time
- **Integration of the WIPCast map display with Google Earth**, complementing the existing Phase I integration with Google Maps
- **Browser-based client,** as an alternative to the existing desktop client, for added ease of deployment and accessibility

5.3.11 SOFTWARE ARCHITECTURE, DESIGN AND SYSTEM INTEGRATION

WIPCast is based on a **Service Oriented Architecture (SOA)**, so as to maximize interoperability with existing systems and frameworks, such as AWRT, T-IWEDA, MyWIDA, DCGS-A, and Google Earth.

Consistent with sound software engineering and architectural principals, the *WIPCast* architecture provides a clean and clear separation between the data layer (i.e., the mesoscale ensemble systems and their output), the logical layer (i.e., *WIPCast* Server) and the user interface layer (i.e., *WIPCast* clients). **Of particular import is the fact that the *WIPCast* system is being architected to avoid reliance on the peculiarities or idiosyncrasies of any particular mesoscale ensemble data source.**

Formal definition of the *WIPCast* Server APIs (Application Programming Interfaces) in Phase II will facilitate integration with other systems such as MyWIDA, T-IWEDA, JMPS, and AWRT. These systems will thereby be able to exploit outputs of the *WIPCast* system, as well as providing mission configuration and other inputs to *WIPCast*.

Specific integration tasks of potential interest in Phase II include:

- Integration and exploitation of AFWA ensemble data
- Acquisition and integration of NWS fine scale analysis grids
- Integration with other systems and services to provide mission profile data (flight route, etc.)
- MyWIDA integration
- Integration with other tools (e.g., Palantier, ParaView, VTK, etc.)

In addition, as discussed in Section 5.3.10, the *WIPCast* front-end geospatial display, which was integrated with GoogleMaps in Phase I, will also be integrated with GoogleEarth in Phase II.

5.3.12 DOCUMENTATION

A comprehensive set of user and technical documentation will be developed as part of the Phase II effort including:

- User Manual
- System Architecture Document

[5] Although a 3D display would appear conceptually to be quite useful, we found it particularly challenging in Phase I to devise a 3D visualization and interface paradigm that was sufficiently intuitive and useful so as to add value beyond our multiple integrated 2D *WIPCast* displays. We nevertheless remain hopeful that some form of 3D display will in fact prove useful as part of the *WIPCast* user interface and therefore intend to explore this area further in Phase II.

- Complete set of JavaDoc for Java classes

Monthly status reports shall also be provided throughout the Phase II period of performance as all as a final technical report upon completion of Phase II.

5.3.13 TESTING

A suite of automated functional and regression test scripts will be developed, maintained and executed as part of the Phase II effort. JUnit (http://www.junit.org/) will most likely be used for automated functional testing and JMeter (http://jakarta.apache.org/jmeter/) will most likely be used for performance testing and benchmarking.

A Verification and Validation (V&V) plan for the system is also to be developed as part of the Phase II effort.

5.3.14 ADDITIONAL RESEARCH AREAS TO CONSIDER

The tasks and research areas detailed in the preceding sections are considered to be the highest priority next steps and have therefore collectively been selected as the focus of the Phase II effort.

There are undoubtedly additional research areas worth pursuing, either in Phase II if time permits, or subsequently as part of further product development and enhancement efforts.

Two particular research topics that bear special consideration are:

- ***Calculating "overall WIP" for the entire mission.*** In the context of decision support, it would be beneficial to provide the decision maker with a "bottom line" overall risk evaluation. Research will need to be conducted to determine the most effective way(s) of synthesizing an overall WIP for the mission, spanning all relevant forecast data across the mission's space-time continuum (e.g., across the entire flight path).
- ***Extending risk evaluation to incorporate other anticipated impacts.*** Using the Army aviation use case as an example, weather impact analysis should not be limited to the flight route when airborne but also should incorporate terminal forecasts since takeoff and landing are often the riskiest parts of the mission.

5.4 PHASE II TEAM

We are pleased to propose a Phase II Project Team that includes all individuals who played a key role in the Phase I effort, in the same roles, thus providing the Army Research Lab (ARL) with the benefit of our team's accumulated, collective project expertise to date.

In particular, the Phase II effort will again be spearheaded by **Impact Computing**, with Hyam Singer serving as Principal Investigator, supported primarily by Mr. Zev Hochberg and Dr. David Makovoz. Extensive support will again be provided by the **University of Washington's Department of Atmospheric Sciences**, under the leadership of Dr. Cliff Mass, supported by Mr. Rick Steed and Mr. Jeff Baars. **Dr. Tony Eckel,** whose seminal work in the domain of probabilistic forecasting and its practical application serves as the foundation for much of our *WIPCast* solution, is to continue as a close and active advisor to our team throughout the Phase II effort.

In addition, the Phase II team will be augmented by Marcus Weather Inc. (spearheading commercialization efforts) and ZedX Inc. (providing hosting services and development support). The presidents of both of these companies – Mr. Kevin Marcus of Marcus Weather and Dr. Joe Russo of ZedX – are to be active members of the Phase II team.

Brief bios of those Phase II team members who served actively on the Phase I team are provided in Section 3.5. Brief bios for those additional individuals who are to augment the Phase I team in Phase II are provided below.

DR. DAVID MAKOVOZ, IMPACT COMPUTING CORP., SYSTEM ARCHITECT

Dr. Makovoz possesses over 25 years of experience in applied R&D. With a background firmly rooted in physical science, Dr. Makovoz has spearheaded a number of innovative, leading edge initiatives across a wide array of complex domains including machine learning, pattern recognition, image processing, signal processing, information retrieval, data mining, modeling and simulation, and natural language processing. Dr. Makovoz possesses a unique blend of experience, spanning algorithm development, mathematical modeling, statistical analysis, complex system architecture and design, and programming in a variety of languages. He possesses expertise in high performance parallel computing and distributed programming on heterogeneous networks. Dr. Makovoz currently serves as Impact Computing's Director of Research and Development. Prior to joining Impact Computing, he served as Chief Engineer at Caltech's Infrared Processing and Analysis Center and Principal Analyst at CACI. Dr. Makovoz possesses a Ph.D. in Physics from the University of Washington as well as an active Secret security clearance.

MR. JOE RUSSO, ZEDX INC., SENIOR TECHNICAL CONSULTANT

Dr. Russo is a co-founder and president of ZedX, Inc., an information technology company that specializes in custom weather databases, decision-support algorithms, and data visualization tools for the agricultural and environmental industries. Since 1999, Dr. Russo has been involved in the design and development of interactive, web-based, decision-support and learning services for the agricultural sector. Since 2002, Dr. Russo has also been involved in the design and development of information technology (IT) platforms for government and university. Dr. Russo has a Ph.D. in Agricultural Meteorology from Cornell University, Ithaca, New York, M.Sc. in Meteorology from McGill University, and a B.S. in Meteorology from St. Louis University. Dr. Russo has published over 25 refereed articles in professional journals, and has been a columnist for Precision Ag magazine since 2006.

MR. KEVIN MARCUS, MARCUS WEATHER INC., METEOROLOGICAL CONSULTANT

Mr. Marcus, CEO and founder of Marcus Weather, is a meteorological consultant who has positively impacted the revenues of numerous weather-sensitive businesses by transforming weather data and forecasts into easy-to-understand, customized reports for key decision makers. He has developed a rule-based weather decision tool for commodity traders that has revolutionized the way that industry uses weather data in their decision making process. Mr. Marcus' career spans more than 30 years, primarily serving clients in the energy and agricultural commodity industries. His meteorological expertise combines with his common sense approach to problem solving to make him a valuable asset to the initiative proposed herein, particularly with an eye toward commercialization. Mr. Marcus possesses an M.S. in Agronomy from the University of Maryland and a B.S. in Meteorology from Penn State University.

5.5 PHASE II SCHEDULE AND MILESTONES

A high-level overview of our planned Phase II schedule and milestones is provided in Figure 18. .

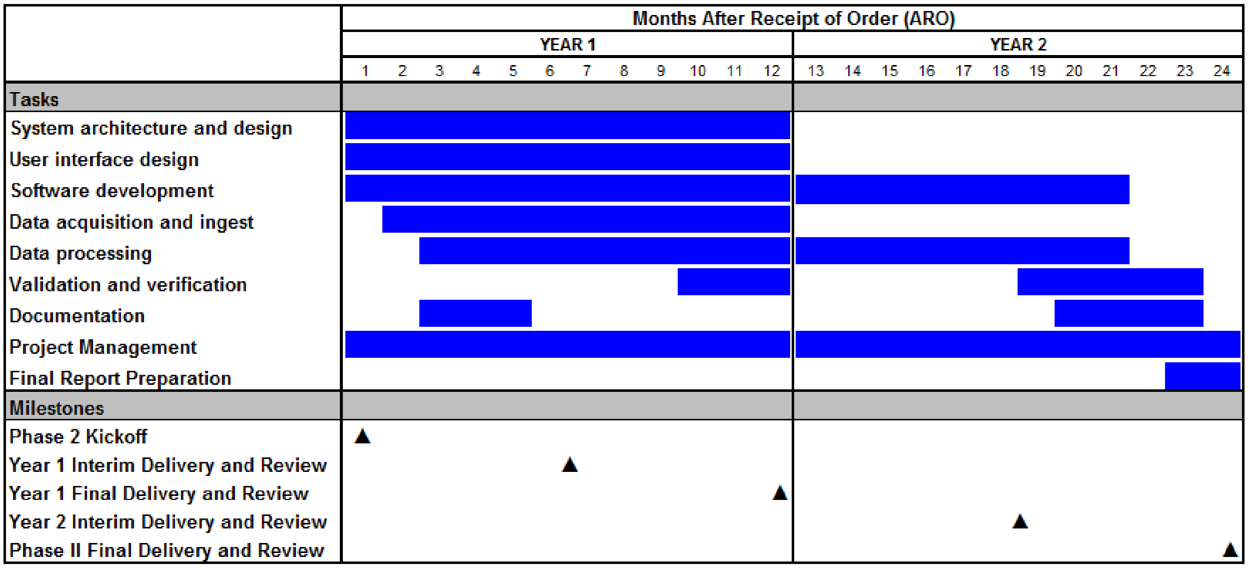


Figure 18. Overview of Planned Phase II Schedule and Milestones

5.6 COMMERCIALIZATION

The *WIPCast* technology proposed herein is expected to garner strong interest across a wide array of Government agencies and weather-sensitive businesses, given the ability of resulting analytical tools to quantify and better anticipate weather-related risk.

Specific industries to be targeted include:

- Commodity traders would benefit greatly from the ability to design a trading program around the probability, and anticipated severity, of weather impacts (similar to mission planning); similarly, hedge funds could use such a tool to exploit anticipated outlier crop production conditions
- Energy companies and utilities could employ WIPCast to support applications that manage and plan for electric demand and generation requirements
- Food manufacturers would benefit from the ability to better plan for price risk on the raw materials
- For the travel industry, probabilistic weather impact analysis would be a vital asset for managing travel risk to determine the best route or time window of travel

Additional anticipated commercial markets and applications include, but are by no means limited to, insurance, transportation, and agriculture.

Probabilistic forecasts are an idea whose time has come. In the energy industry, for example, it is estimated that hundreds of millions of dollars could be saved annually through use of improved probabilistic forecasts vs. the current deterministic forecasts used to forecast wind generation, energy demand and allocation of resources. One example of where improved forecasts for energy demand can be gained is through probabilistic weather forecast inputs into the load models used to predict energy demand across the electric grid system. Deterministic forecasts are routinely used that include no quantitative measure of inherent risk in the forecast of temperature, humidity, sky cover and winds. A more robust approach would be to construct an ensemble of weather inputs that will quantify risk by hour. Such an application for improved short term forecasts would be worth millions of dollars to the electric energy industry.

Similarly, commodity trading and purchasing strategies for hedge funds and food manufacturers, as well as electric demand/generation applications for energy utilities, could generate significant *WIPCast* subscription revenue in the first 5 years alone. Also for the travel industry, avoiding poor or unsafe travel conditions translates to millions of dollars of revenue each year. The leisure, commercial trucking, and private air travel industries spend $5-10MM per year for basic weather information. A decision support system that combines travel risk tolerance with uncertainty in weather forecasts could itself be a several million dollar per year market.

Along the same lines, agriculture businesses spend tens of millions of dollars for weather data and forecasts. A decision-based product already provided for commodity traders and food manufacturers commands two to three times the industry standard for private weather services. Such multiples can be applied across a number of weather sensitive industries once the *WIPCast* algorithms are adapted to the context of the economic impacts that affect these other businesses.

Marcus Weather Inc., our Phase II partner, is well positioned to lead the market penetration, since it has already established a beach head in these industries through 30 years of providing service, developing new weather based interpretative products, adapting new technology for rapid assessment of short term and longer term weather risk.

With regard to licensing and pricing, we at Impact Computing are strong proponents of the open source software licensing model. Consistent with this philosophy, *WIPCast* is to be licensed and distributed as open source software. It is important to note, though, that our belief in the open source model is not, strictly philosophical, it is also practical. Open source has clear benefits from the standpoint of market penetration. Companies that offer open source software are often able to establish an industry standard and thereby gain competitive advantage. Many companies have also benefited from the fact that the open source model helps build developer loyalty as developers feel empowered and have a sense of ownership of the end product. Moreover, the costs of marketing, packaging and distributing open source technologies are typically less – often far less – than those costs for more traditionally protected and licensed competitive technologies.

In fact, recent years have witnessed explosive growth in the success of open source technologies in attaining dominant market share positions. Examples include the Apache Web Server, the Red Hat distribution of the Linux operating system, the MySQL relational database, the Mozilla Firefox web browser, and the Sendmail email server. Each has demonstrated sufficient stability, capability and maturity to capture significant, and in some cases dominant, market share, even against alternative products from commercial software vendors no less formidable than Microsoft and Oracle.

Beyond the benefits of open source from the standpoint of market penetration, it is also important to recognize that open source products can and do produce revenue streams. Red Hat and MySQL are prime examples of companies that have built viable, highly successful, commercial businesses based on open source products and the open source revenue model, whereby revenue streams are primarily the result of product-related subscriptions and custom services. *WIPCast*-related revenue is therefore expected to be generated from these sources as well.

WIPCast development beyond the initiative proposed herein is expected to be accomplished through a combination of open source contributions, internal investment, and funded enhancements as summarized in Figure 19. .

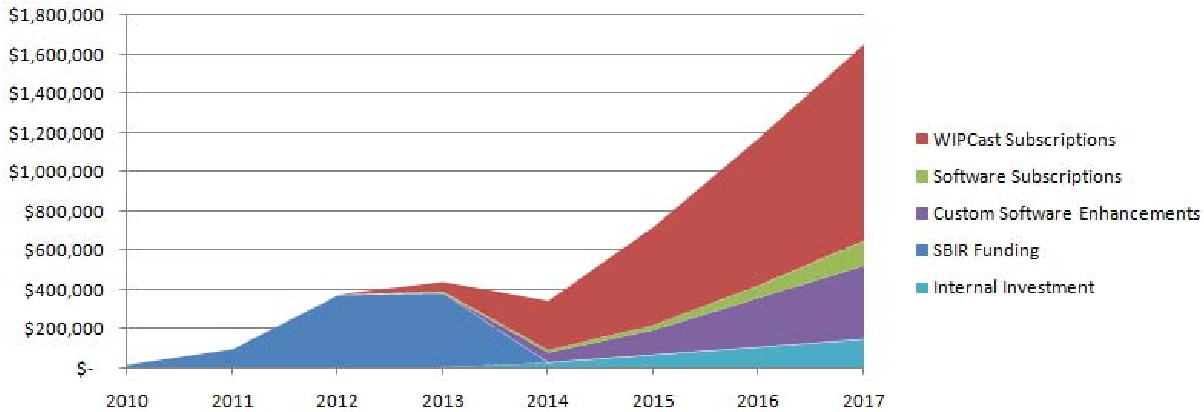


Figure 19. Projected WIPCast-related Revenues and Investment

6 CLOSING REMARKS

The combined Impact Computing / University of Washington team has greatly appreciated the opportunity to support the Government on this important groundbreaking project in the domain of decision support tools that exploit the power of probabilistic forecasting.

Our recently completed Phase I initiative has been both challenging and rewarding and we look forward to hopefully proceeding to a follow-on Phase II effort.

REFERENCES

Allen, M. S., 2009: Ambiguity in Ensemble Forecasting: Evolution, Estimate Validation, and Value. Ph.D. Dissertation, Graduate School of Engineering and Applied Sciences, Naval Postgraduate School, [Available from the Defense Technical Information Center at http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA509168&Location=U2&doc=GetTRDoc.pdf]

-----, and F. A. Eckel, 2011: Value from ambiguity in ensemble forecasts. *Wea. Forecasting*. In Review [Available from tony.eckel@noaa.gov]

Air Weather Service, 1991: AWS/TR--91/001, Probability Forecasting; A Guide for Forecasters and Staff Weather Officers. United States Air Force, Air Weather Service, Scott Air Force Base, IL. (Originally published in 1978 as AWS PAMPLET 105-51)

Boyle, J. S., 1990: Turbulence indices derived from FNOC field and TOVS retrievals: Naval Oceanographic and Atmospheric Research Laboratory, Technical Note 47; 22 pp.

Brooks, G. R. and A. D. Oder, 2004?: Low-level turbulence algorithm testing at-or-below 10,000 ft. *AMS Conference Preprint.* Available at [ams.confex.com/ams/pdfpapers/82122.pdf]

Berrocal, V.J., Raftery, A.E. and Gneiting, T., 2007: Combining Spatial Statistical and Ensemble Information in Probabilistic Weather Forecasts. Monthly Weather Review, 135, 1386-1402.

-----, Raftery, A.E. and Gneiting, T., 2008: Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. Annals of Applied Statistics 2: 1170-1193.

-----, Raftery, A.E. and Gneiting, T. Probabilistic Weather Forecasting for Winter Road Maintenance. Journal of the American Statistical Association, in press.

Chmielecki, R.M. and Raftery, A.E., 2010: Probabilistic visibility forecasting using Bayesian Model Averaging. *Monthly Weather Review* ,139, 1626--1636.

Eckel, F. A., 2003: Effective mesoscale, short-range ensemble forecasting. Ph.D. dissertation, University of Washington, 224 pp. [Available online at http://www.atmos.washington.edu/~ens/pubs_n_pres.html]

----- and C.F. Mass, 2005: Aspects of effective mesoscale short-range ensemble forecasting. *Wea. Forecasting*, 20, 328-350.

-----, Cunningham J.G., and D.E. Hetke, 2008: Weather and the calculated risk: Exploiting forecast uncertainty for Operational Risk Management. Air & Space Power Journal, 22, 71–82.

-----, M.S. Allen, and M.C. Sittel, 2011: Estimating ambiguity in ensemble forecasts. *Wea. Forecasting*. In Review [Available from tony.eckel@noaa.gov]

Ellrod, G.P., and D.I. Knapp, 1992: An objective clear-air turbulence forecasting technique: Verification and operational use. *Wea. Forecasting*, **7**, 150-165.

Gel, Y., Raftery, A.E. and Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with Discussion). Journal of the American Statistical Association, 99, 575-590.

Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecasts. *Mon.Wea. Rev.*, **125**, 1312–1327.

National Research Council Committee on Estimating and Communicating Uncertainty in Weather and Climate Forecasts, 2006: Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts. The National Academies Press, Washington, D.C., 112 pp. [Available at http://www.nap.edu/catalog/11699.html]

Passner, J., 2000: An evaluation of the three-dimensional weather hazards using sounding data and model output data. ARL-TR-1046.

Raftery, A.E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005). Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Monthly Weather Review*, 133, 1155-1174.

Shirkey, R.C. and Gouveia, M., 2002: Weather-impact decision aids: Software to help plan optimal sensor and system performance. CrossTalk: The Journal of Defense Software Engineering, December 2002

Szczes, J.R., 2008: Communicating Optimized Decision Input from Stochastic Turbulence Forecasts. M.S. Thesis, Graduate School of Engineering and Applied Sciences, Naval Postgraduate School 159 pp. [Available from the Defense Technical Information Center at http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA480037&Location=U2&doc=GetTRDoc.pdf]